AD-761 792

ELECTRO-THERMAL OVERSTRESS FAILURE IN MICROELECTRONICS

CLARKSON COLLEGE OF TECHNOLOGY

PREPARED FOR
ROME AIR DEVELOPMENT CENTER

APRIL 1973

RADC-TR- 73-87
Final Technical Report
April 1973

ELECTRO-THERMAL OVERSTRESS FAILURE IN MICROELECTRONICS

Clarkson College of Technology

**Approved for public release;
distribution unlimited.**

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Clarkson College of Technology Potsdam, New York	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b N/A

3 REPORT TITLE

ELECTRO-THERMAL OVERSTRESS FAILURE IN MICROELECTRONICS

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Final Report

5 AUTHOR(S) (First name, middle initial, last name)

Dr. Henry Domingos

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
April 1973	182 194	32

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
F30602-72-C-0007 Job Order No. 55190629	None 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-73-87

10 DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
None	Rome Air Development Center (RBRP) Griffiss Air Force Base, New York 13441

13 ABSTRACT

Failure of microelectronic circuits caused by electrical overstress has been investigated theoretically and experimentally. Computer calculations on heat flow in silicon structures have pointed out the necessity of using temperature dependent thermal constants for silicon. The effect of surface layers and power distribution on the peak temperature in silicon devices under transient conditions has been investigated.

Integrated circuits representing several fabrication technologies from several different manufacturers were pulsed to failure over pulse lengths from 100 nsec to 10 msec. The threshold power per unit area to produce permanent damage was found to vary by a factor of 20 among the parts tested. Peak temperatures to initiate current constriction were estimated to be 1000-1200°C.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Reliability						
Solid State Devices						
Breakdown						
Silicon						

Security Classification

ELECTRO-THERMAL OVERSTRESS FAILURE IN MICROELECTRONICS

Dr. Henry Domingos

Clarkson College of Technology

FOREWORD

This Final Report describes research by Clarkson College of Technology, Potsdam, New York, under contract F30602-72-C-0007, Job Order Number 55190629, for Rome Air Development Center, Griffiss Air Force Base, New York. Mr. John F. Carroll (RBRP) was the RADC Project Engineer.

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service.

This technical report has been reviewed and is approved.

Approved: JOHN F. CARROLL
Reliability Physics Section
Reliability Branch

Approved: JOSEPH J. NARESKY
Chief, Reliability & Compatibility Division

FOR THE COMMANDER:

CARLO P. CROCETTI
Chief, Plans Office

ABSTRACT

Failure of microelectronic circuits caused by electrical overstress has been investigated theoretically and experimentally. Computer calculations on heat flow in silicon structures have pointed out the necessity of using temperature dependent thermal constants for silicon. The effect of surface layers and power distribution on the peak temperature in silicon devices under transient conditions has been investigated.

Integrated circuits representing several fabrication technologies from several different manufacturers were pulsed to failure over pulse lengths from 100 nsec to 10 msec. The threshold power per unit area to produce permanent damage was found to vary by a factor of 20 among the parts tested. Peak temperatures to initiate current constriction were estimated to be 1000-1200°C.

EVALUATION

The objective of this effort was to provide a physical basis for assessing the reliability of microelectronic devices when subjected to electrical overstress. The contractor performed a combined analytical and experimental program to define the significant parameters associated with overstress failure of microelectronic structures. While considerable overstress testing has been done on microelectronic structures, this effort adds substantial understanding to the detailed mechanisms of failure. This was done by the use of well controlled experimental procedures and the thorough analysis of failed devices and associated electrical waveforms. Dr. H. Domingos was able to relate his results on emitter base flashover shorts in integrated circuits to recent work on diode second breakdown in thin silicon film on sapphire. This provides a good physical understanding of the surface flashover type of breakdown in microelectronic structures. In addition, he demonstrated that differences in ability to withstand overstress in the circuits tested were related to the distance from the P-N junction to the contact on the high resistivity side of the junction and the existence of multiple conduction paths in the microelectronic device.

The information obtained from this effort will be used in the interpretation of failures due to overstress in microelectronic systems. Additional in-house effort in the electrical overstress area is being considered using test structures consistent with LSI processing and LSI circuits.

John F. Carroll

JOHN F. CARROLL
Reliability Physics Section
Reliability Branch

TABLE OF CONTENTS

Section		Page
I.	Introduction	1
II.	Heat Transfer in Integrated Circuits	3
	A. General Heat Flow	3
	B. Heat Flow by Conduction	4
	C. Thermal Conductivity of Silicon	20
	D. Numerical Solutions to Heat Flow Problems	26
III.	A Survey of Second Breakdown in Integrated Circuits	47
IV.	Experimental Results	59
	A. Testing Procedure	59
	B. Devices Tested	62
	C. Results	62
V.	Discussion	157
VI.	Conclusion	168
	References	171
	Appendix	174

LIST OF ILLUSTRATIONS

Figure		Page
1.	Temperature distribution in a semi-infinite solid due to a power density F_o watts/cm^2 into the surface.	8
2.	Temperature rise at the surface of a semi-infinite solid due to a power density F_o watts/cm^2 into the surface.	9
3.	Temperature distribution due to a constant power density F_o watts/cm^2 into the surface of a slab terminated at x = L with a perfect heat sink.	10
4.	Temperature rise at the surface of a slab terminated at x = L with a perfect heat sink.	11
5.	Temperature distribution in a semi-infinite solid due to a power density P_o watts/cm^3 in a layer of thickness L at the surface.	13
6.	Temperature rise at the surface of a semi-infinite solid with a power input P_o watts/cm^3 in a layer of thickness L at the surface.	14
7.	Temperature distribution in a semi-infinite body due to a constant power applied at the surface over a circular area of radius r_o.	18
8.	Temperature rise in a semi-infinite solid due to power applied at the surface over a circular area of radius r_o.	19
9.	Thermal conductivity of pure silicon as a function of absolute temperature.	23
10.	A graphical solution for the temperature rise in an integrated circuit with two different approximations for the thermal conductivity.	25
11.	Analogy between the thermal model (top) and the electrical model (bottom).	28
12.	Temperature distribution in a silicon diode, solved by an explicit, finite difference method.	30
13.	Temperature distribution in silicon due to power applied in a 4 μm layer at the surface.	36

Figure		Page
14.	Temperature distribution · der the same conditions as Fig. 13, but with a 0.5 μm layer of silicon dioxide and a 1 μm layer of aluminum at the surface.	37
15.	Effect of surface layers on peak temperature rise in silicon.	39
16.	Peak temperature rise in silicon after 1.5 msec with variable thicknesses of silicon dioxide and aluminum.	41
17.	Power versus time curves to reach the maximum temperature indicated.	42
18.	Same as Fig. 17, with the power applied to a 3 μm layer.	43
19.	Same as Fig. 17, with the power applied to a 4 μm layer.	44
20.	Second breakdown characteristics of semiconductor devices	49
21.	Pulse voltage and current waveforms of a device stressed into second breakdown.	50
22.	Experimental arrangement for pulsing devices.	61
23.	Pin connections and schematic diagram for Part A.	64
24.	Part A, dual 3-input gate expander.	66
25.	Cross-section view of Part A.	68
26.	Pulse waveforms for Part A in the reverse bias mode.	70
27.	Threshold power as a function of pulse width for reverse-biased emitter junctions of Part A.	74
28.	Forward bias waveforms for Part A.	75
29.	Threshold power vs time curves for Part A.	79
30.	Reverse bias breakdown characteristics showing damage due to second breakdown.	81
31.	Damage due to reverse bias pulses.	82
32.	Search for junction damage.	85
33.	Damage to forward biased junctions.	88

Figure		Page
34.	Schematic diagram for Part B.	91
35.	Part B, special test circuit.	92
36.	Cross-section view of Part B.	94
37.	Electrical characteristics, Part B.	96
38.	Reverse bias waveforms for the diode of Part B.	99
39.	Forward bias waveforms for the diode of Part B.	101
40.	Threshold power vs time curves for the diode of Part B.	102
41.	Diode damage due to reverse bias pulses.	104
42.	Emitter-base pulse waveforms for Part B, reverse bias.	107
43.	Power versus delay time curves for Part B emitter-base junctions.	111
44.	Emitter-base junction damage.	112
45.	Waveforms for collector-base junctions, Part B.	114
46.	Damage due to collector-base pulses.	117
47.	Threshold damage for collector-base junctions, Device B.	120
48.	Collector-emitter breakdown waveforms for transistors of Device B.	122
49.	Contact damage produced by pulses at/or just below first transition levels.	126
50.	Collector-emitter breakdown waveforms showing the second transition.	128
51.	Contact damage produced by pulses with second transition levels.	129
52.	Threshold damage curve for collector-emitter pulsing, Part B.	132
53.	Photomicrographs of the circuit of Part C.	135
54.	Pin connections for the input diodes and details of the construction.	136

Figure		Page
55.	Diode characteristics.	138
56.	Breakdown waveforms.	139
57.	Threshold damage curves for the three diode configurations of Device C.	142
58.	Part D input diodes.	144
59.	Breakdown waveforms.	146
60.	Threshold damage curve for Part D.	148
61.	One of the input diode groups on Part E.	149
62.	Breakdown waveforms for the diodes of Part E.	150
63.	Threshold power curve for Device E.	152
64.	Breakdown waveforms for core driver transistors.	154
65.	Threshold damage points for the core driver transistors of Part F.	156
66.	Current, Power, and temperature distribution in Part A for the reverse bias case.	161

LIST OF TABLES

Table		Page
I.	Test Devices	63
II.	Diode Breakdown Calculations from Figure 26(c)	72
III.	Threshold Power and Power Per Unit Area	164

SECTION I

INTRODUCTION

It is a well known fact that voltage, current, and power ratings of semiconductor devices must be scrupulously observed in any application. Unlike vacuum tubes, where operation above rated values merely reduces the operating lifetime, overstress of semiconductor devices often leads to catastrophic failure. Because electronic circuits must function in environments where voltage and current surges due to power supply transients, line disturbances, electromagnetic pulses, etc., may be present, a clear understanding of failure due to such conditions is needed. Apart from the uneasiness which lack of understanding generates, it can also result in systems which are either unreliable or needlessly overdesigned. Better understanding leads to better component design as well as more efficient engineering and more reliable equipment.

The type of failure investigated in this study is degradation due to extreme temperature rises resulting from excessive power dissipation as distinct from degradation due to such mechanisms as aluminum migration, purple plague, chemical and metallurgical reactions, etc. Indeed, in any device, semiconductor or otherwise, excessive power dissipation is of consequence only because it raises the temperature: as long as the device can be kept cool the power dissipation is a secondary consideration. In semiconductor devices the temperature rise eventually leads to melting of the silicon or other materials comprising the circuit.

The earliest efforts to systematically study device failure were undertaken in the late fifties. The recognition of the phenomenon known as second breakdown caused a flurry of activity in the mid-sixties. This has

not yet abated as the development of high pulse power microwave diodes has entered the scene.

Work done at other laboratories with silicon-on-sapphire diodes while this program was in progress has led to a much clearer understanding of second breakdown in semiconductor devices, particularly single junctions. The results of the present study tend to agree with these findings and thus it can be said that a good picture of second breakdown in diodes is now available.

This report is the final report on a program of theoretical and experimental investigations on breakdown in microelectronic circuits. First a study of heat flow in silicon is presented If the spatial distribution of the dissipated power is known, a knowledge of the thermal properties of the materials enables the temperature distribution to be calculated. Unfortunately the power dissipation depends on the characteristics of the source as well as the electrical characteristics of the circuit, which in turn depends on the temperature. This leads to a complex situation. Finite-difference solutions have been obtained for temperature distributions in idealized one-dimensional situations taking into account the variation of thermal parameters with temperature.

A section on review of second breakdown theories is followed by a presentation of the experimental results. Several types of microelectronic circuits were pulsed into second breakdown to determine threshold energies for damage as a function of pulse width and to determine the nature of the damage. The results are discussed in terms of current models of second breakdown and the heat flow calculations.

SECTION II

HEAT TRANSFER IN INTEGRATED CIRCUITS

A: General Heat Flow

When power is applied to a silicon monolithic integrated circuit, most of the energy is dissipated in the circuit elements within the chip. The heat can be carried away by conduction, convection, and radiation to other parts of the chip and ultimately to the package and to the ambient. Nearly all of the heat flow is through the silicon to the header and hence to the rest of the package. Heat can also flow along metallization runs to the leads and then to the package; through the oxide and the ambient gas in the package; or be radiated to cooler parts of the structure. Heat transfer by all three processes, radiation, convection, and conduction was investigated; only transfer by conduction was found to be significant in typical microcircuits.

Convective heat transfer is sometimes interpreted to also include conduction by the convecting medium; however, it differs from pure conduction in that the transmission of energy is augmented by movement of the fluid itself. Free convection comes about as a result of the change in buoyancy of the fluid as a result of uneven heating. If the fluid is confined to a very small space the buoyant forces are not sufficient to overcome viscosity, and fluid motion, hence convection, does not occur. This is the case for air inside an integrated circuit package.

Radiation is electromagnetic in nature and radiative exchanges are governed by the laws of electromagnetic theory or optics. Suppose we consider radiation from the hot surface of a chip to the gold-plated lid of the package. Treating the two surfaces as specularly reflecting plates at 1000°K

and 290°K, respectively, a reasonable value for the heat transfer coefficient is 0.2 watts/cm^2.[1] For a chip 100 mils on a side radiation would account for 12.9 mw. This is a liberal estimate since not all of the chip surface would be at 1000°K, and the thermal oxide or deposited glass is a good insulator. Even so, this power level is about three orders of magnitude less than power levels of concern in pulse testing.

Radiation from the surface down through the silicon to the header is of even less importance. In the first place silicon has a fundamental absorption edge at 1.13 μm. Radiation at shorter wavelengths would be sharply attenuated by the excitation of electrons to the conduction band. The dominant wavelength of a black body radiator at 1000°K is 2.9 μm. Radiation at this wavelength will be attenuated by joule heating of the lossy silicon. For 1 ohm-cm material the penetration depth is only 5 μm.

B. Heat Flow by Conduction

The differential equation which describes the conduction of heat in a solid is basically a continuity equation for the conservation of heat energy and can be written in its most general form as

$$\rho c \frac{\partial T}{\partial t} = q + k\nabla^2 T + \frac{\partial k}{\partial T}\left[\left(\frac{\partial T}{\partial x}\right)^2+\left(\frac{\partial T}{\partial y}\right)^2+\left(\frac{\partial T}{\partial z}\right)^2\right] \qquad (1)$$

where ρ is the density in grams/cm^3, c is the specific heat in joules/gram-°C, T is the temperature in °C, t is in seconds, q is the power density in watts/cm^3, k is the thermal conductivity in watts/cm-°C, and x, y, and z are space coordinates in cm. q can be a function of time and space and k can be a function of temperature.

As written, Equation (1) is non-linear and can be solved only in certain special cases. If k is not a function of temperature, the equation

becomes a linear partial differential equation for which many solutions are tabulated.[2]

Solutions to the heat flow equation must satisfy the initial condition that at t = 0

$$T = f(x,y,z)$$

and one or more boundary conditions

i) temperature at one or more surfaces is constrained

$$T_{surf} = g(x,y,z,t)$$

ii) a surface is insulated so that no heat flows across it

$$\frac{\partial T}{\partial n} = 0$$

iii) miscellaneous boundary conditions

$$k \frac{\partial T}{\partial n} + H(T-T_o) - F = 0 \ .$$

The first term is heat flow across the surface by conduction, the second term is the "radiation" boundary condition and represents a heat loss proportional to the difference between the surface at temperature T and another surface at T_o, and the last term is the heat flux into the surface from an external source.

A number of simple solutions to heat flow problems will now be applied directly to special situations occuring in microcircuits. The following constants for silicon will be used in the calculations that follow:

$\rho = 2.328$	gm/cm^3
$c = 0.7$	$J/gm-^oC$
$k = 1.45$	$W/cm-^oC$
$D = 0.9$	cm^2/sec

1. Adiabatic heat problem

Consider a small volume initially at a uniform temperature T_o to which a uniform power density P_o watts/cm^3 is supplied throughout, with insulated boundaries. The differential equation becomes

$$\frac{\partial T}{\partial t} = \frac{P_o}{\rho c} \tag{2}$$

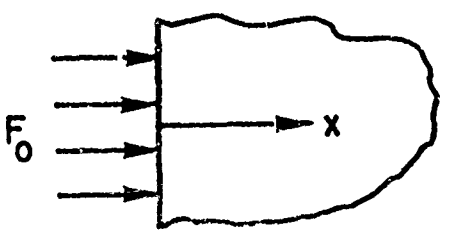
with the solution

$$T - T_o = \frac{P_o t}{\rho c} = 0.614\ P_o t \quad ^\circ C \tag{3}$$

where t is in seconds and P_o is in watts/cm^3. This solution is applicable to thermal problems during a time interval short enough that flow by diffusion is negligible, as, for instance, estimating the temperature rise in a junction during the first few nanoseconds after power is applied. The equation gives an upper limit to the temperature rise and a reasonable accurate answer as long as $\sqrt{Dt}$ is less than the dimensions of the region.

2. Constant power into the surface of a semi-infinite medium

This is a one-dimensional problem where the silicon extends from x = 0 to infinity, with a power density F_o in watts/cm^2 at the surface.



The differential equation is

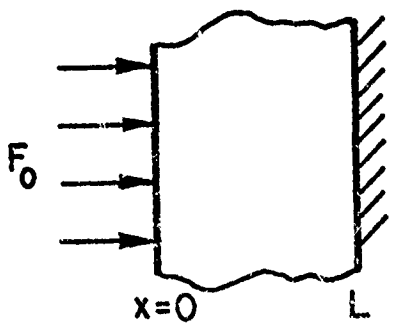
$$\frac{\partial T}{\partial t} = \frac{k}{\rho c} \frac{\partial^2 T}{\partial x^2} = D \frac{\partial^2 T}{\partial x^2} \tag{4}$$

The solution is

$$T-T_o = \frac{2F_o}{k} \left\{ \left(\frac{Dt}{\pi}\right)^{\frac{1}{2}} e^{-x^2/4Dt} - \frac{x}{2} \operatorname{erfc} \frac{x}{2(Dt)^{\frac{1}{2}}} \right\} \qquad (5)$$

This solution is applicable to the case of uniform power dissipation in a comparatively thin layer at the surface of a chip. The solution is valid as long as lateral heat flow and diffusion to the header can be neglected. Equation (5) is plotted in normalized form in Figure 1 and again in Figure 2 as a function of time for x = 0. For a total power of 25 watts into an area 10 mils x 10 mils, $F_o = 40 \times 10^3$ watts/cm^2 and the maximum abscissa in Figure 1 is a temperature rise of 480°C.

3. Constant power into the surface of a semiconductor of finite thickness, with an ideal heat sink



This case applies when the finite thickness of the die must be considered. It is assumed that edge effects are still negligible, and that the header is able to maintain the x = L plane at constant temperature T_o. The solution is

$$T-T_o = \frac{F_o(L-x)}{k} - \frac{8F_oL}{k\pi^2} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} e^{-\frac{D(2n+1)^2\pi^2 t}{4L^2}} \sin \frac{(2n+1)\pi(L-x)}{2L} \qquad (6)$$

This is plotted in Figure 3. Figure 4 shows the temperature rise as a function of time at x = 0. For 25 watts into an area 10 mils x 10 mils

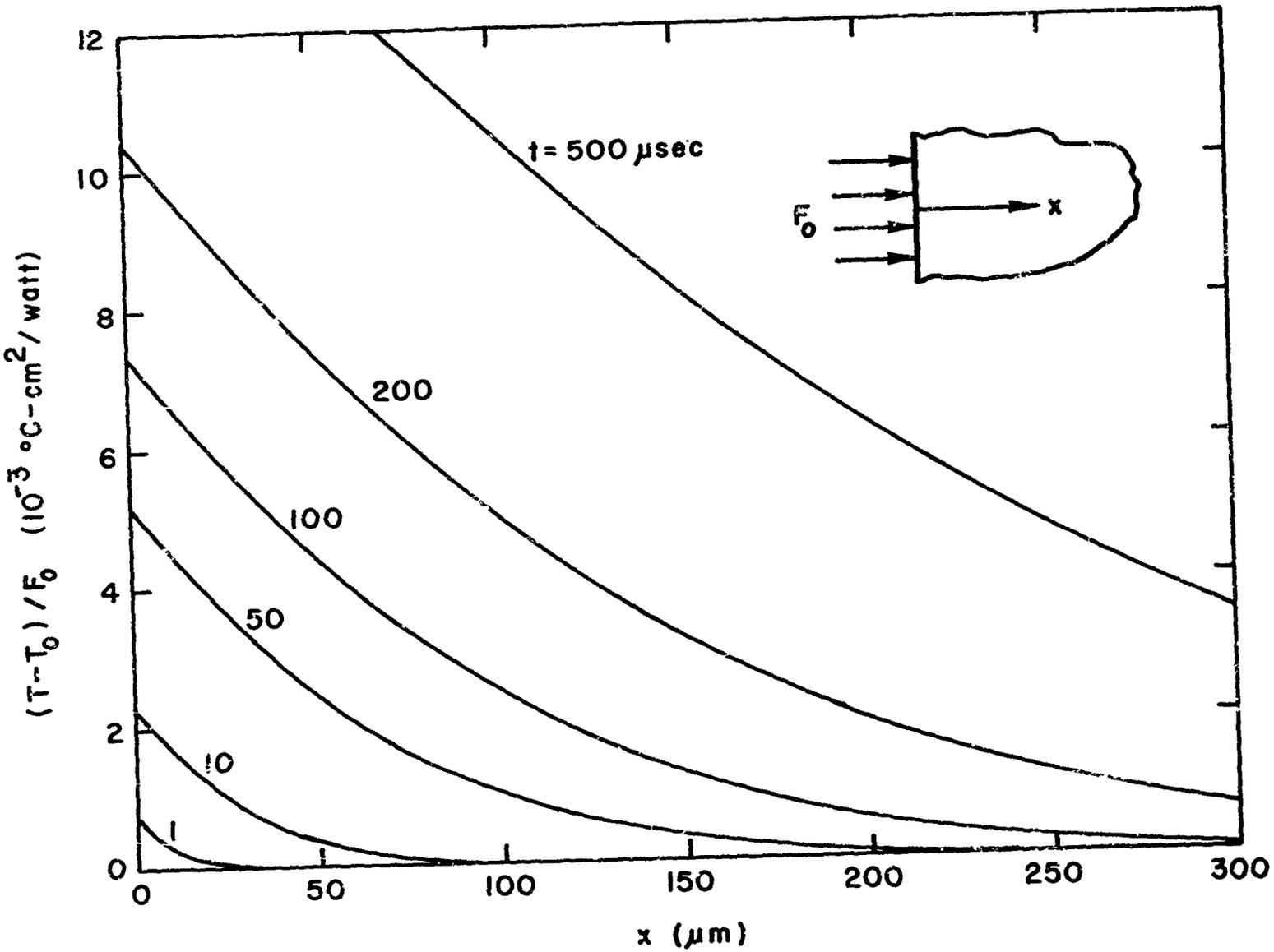


Fig. 1. Temperature distribution in a semi-infinite solid due to a power density F_o watts/cm^2 into the surface.

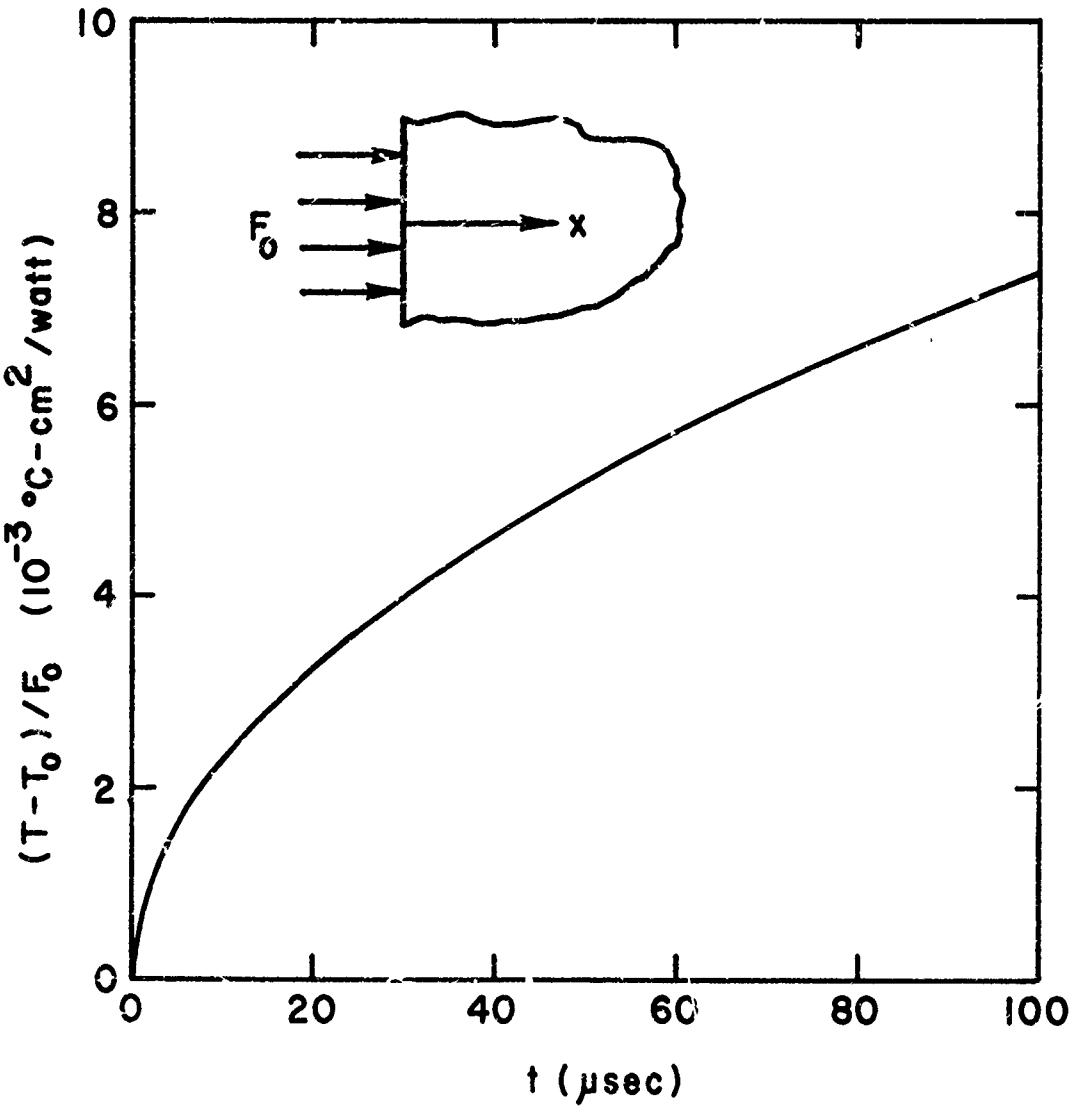


Fig. 2. Temperature rise at the surface of a semi-infinite solid due to a power density F_o watts/cm^2 into the surface.

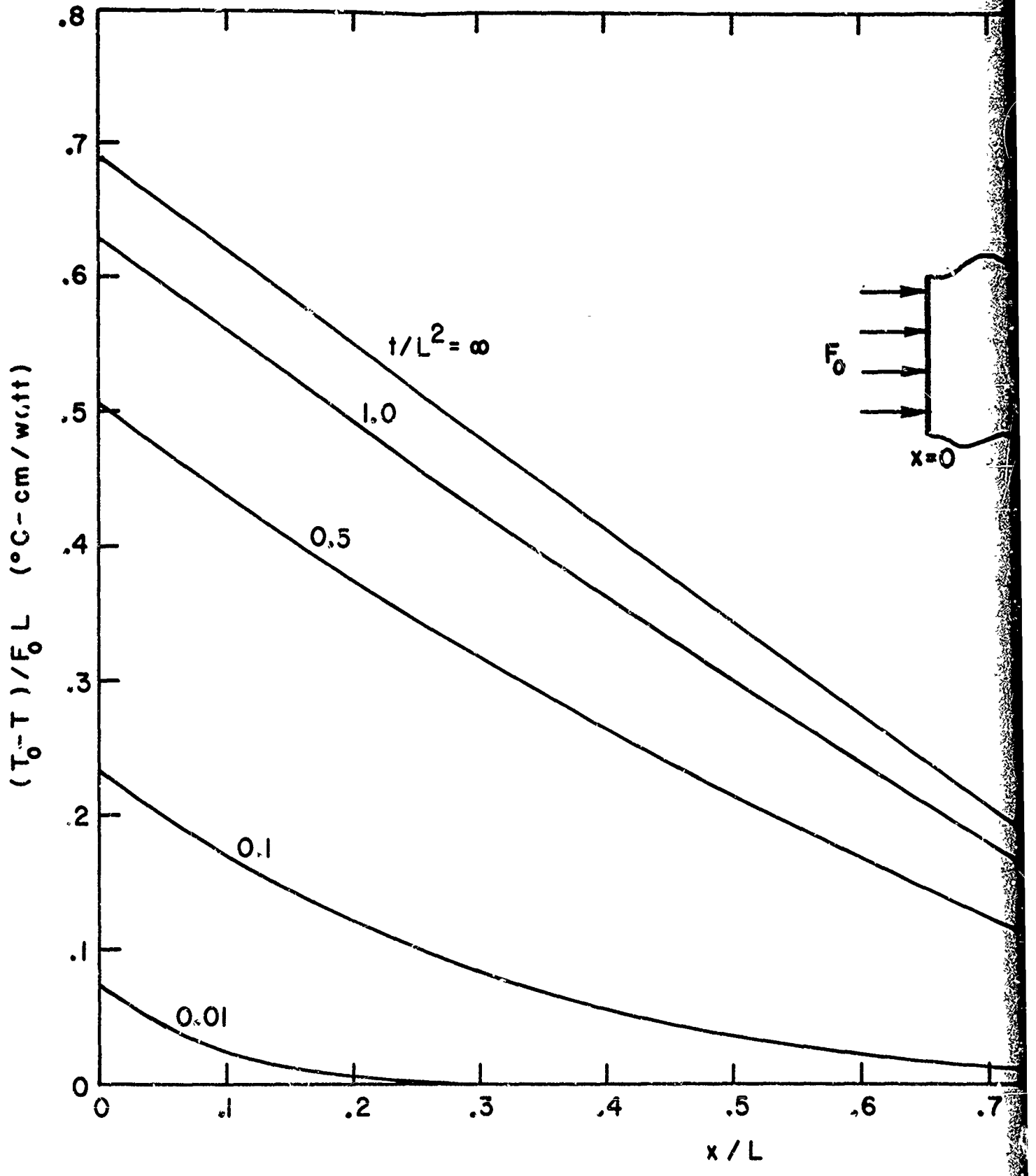
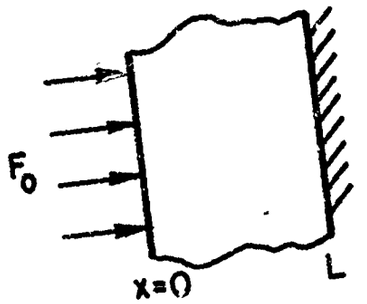


Fig. 3. Temperature distribution due to a constant power de
the surface of a slab terminated at x = L with a pe



.4 .5 .6 .7 .8 .9 1.0

x / L

e distribution due to a constant power density F_o watts/cm² into
e of a slab terminated at x = L with a perfect heat sink.

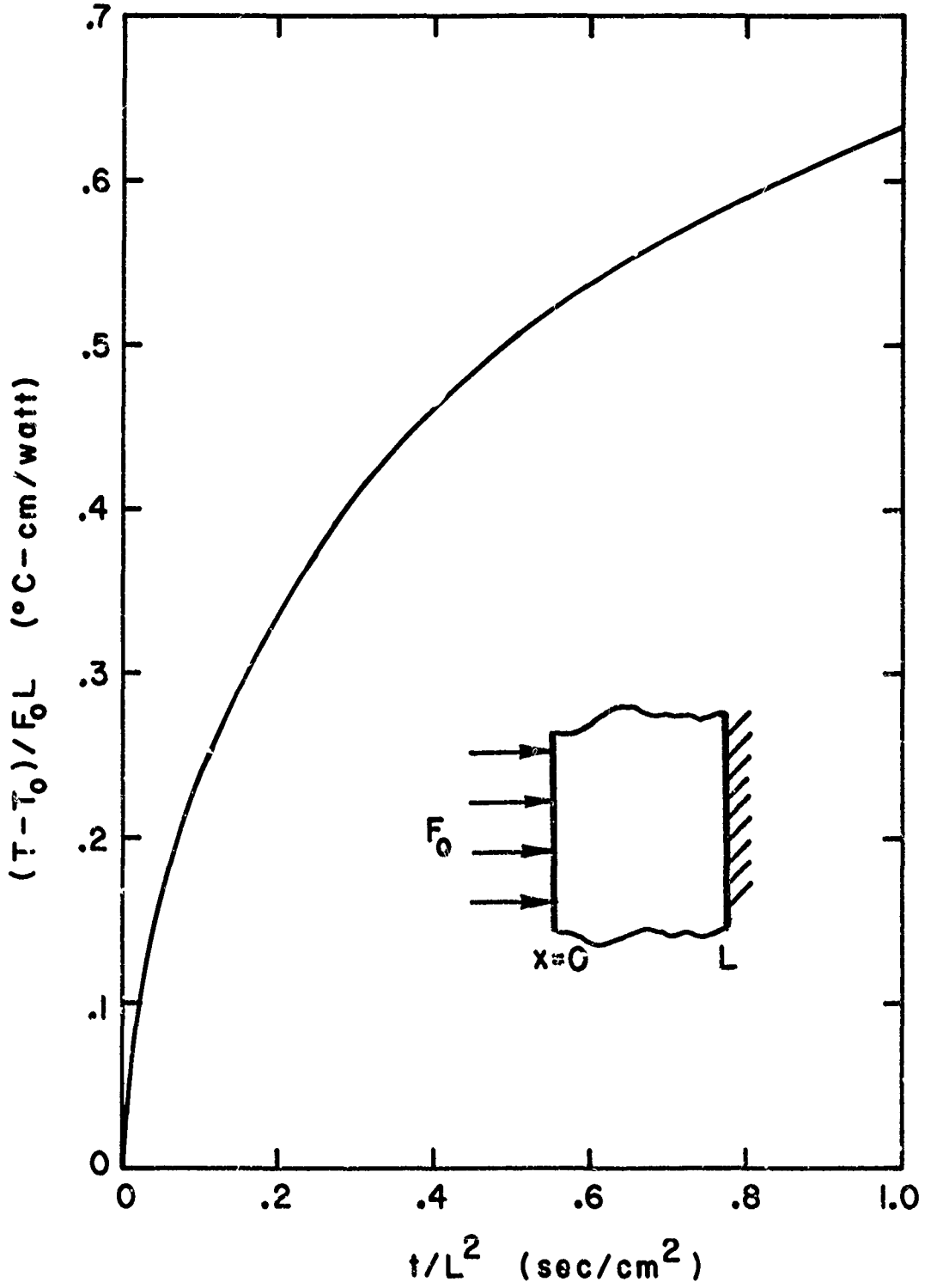
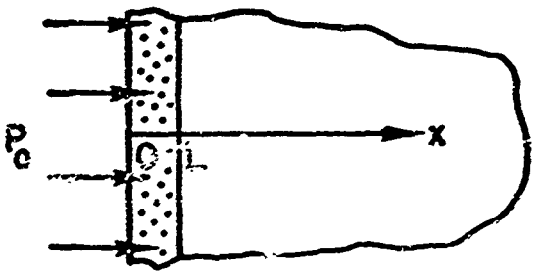


Fig. 4. Temperature rise at the surface of a slab terminated at x = L with a perfect heat sink. The asymptotic value is 0.69.

and with L = 10 mils, multiply the ordinate by 1000. $t/L^2 = 0.1$ then corresponds to 62.5 μsec. The steady state temperature rise in this case is 690°C, which can be verified by calculating the thermal resistance. For the general case, the presence of the heat sink is important only for $t/L^2>0.2$, and steady-state is reached for $t/L^2>10$.

4. Semi-infinite region, with power density P_o watts/cm³ into a finite thickness at the surface



The solution is

$$T-T_o = \frac{P_o t}{\rho c}\{1-2\ i^2\text{erfc}\ \frac{L-x}{2\sqrt{Dt}} - 2\ i^2\text{erfc}\ \frac{L+x}{2\sqrt{Dt}}\} \text{ for } 0<x<L \qquad (7)$$

$$= \frac{2P_o t}{\rho c}\{i^2\text{erfc}\ \frac{x-L}{2\sqrt{Dt}} - i^2\text{erfc}\ \frac{x+L}{2\sqrt{Dt}}\} \text{ for } x>L$$

where $$i^2\text{erfc}\ \mu = \tfrac{1}{4}\{(1+2\mu^2)\text{erfc}\ \mu - \frac{2}{\sqrt{\pi}}\mu e^{-\mu^2}\}$$

This solution is useful when edge effects and finite length effects can be neglected, but when the heat-dissipating source at the surface must be considered in more detail. The thickness of epitaxial layers and depths of diffusion for example are often important parameters. This model is also the basis of numerical calculations presented later in this section.

Equation (7) is plotted in Figures 5 and 6. For comparison, we can again take 25 watts into an area 10 mils x 10 mils with L = 10 μm. In this

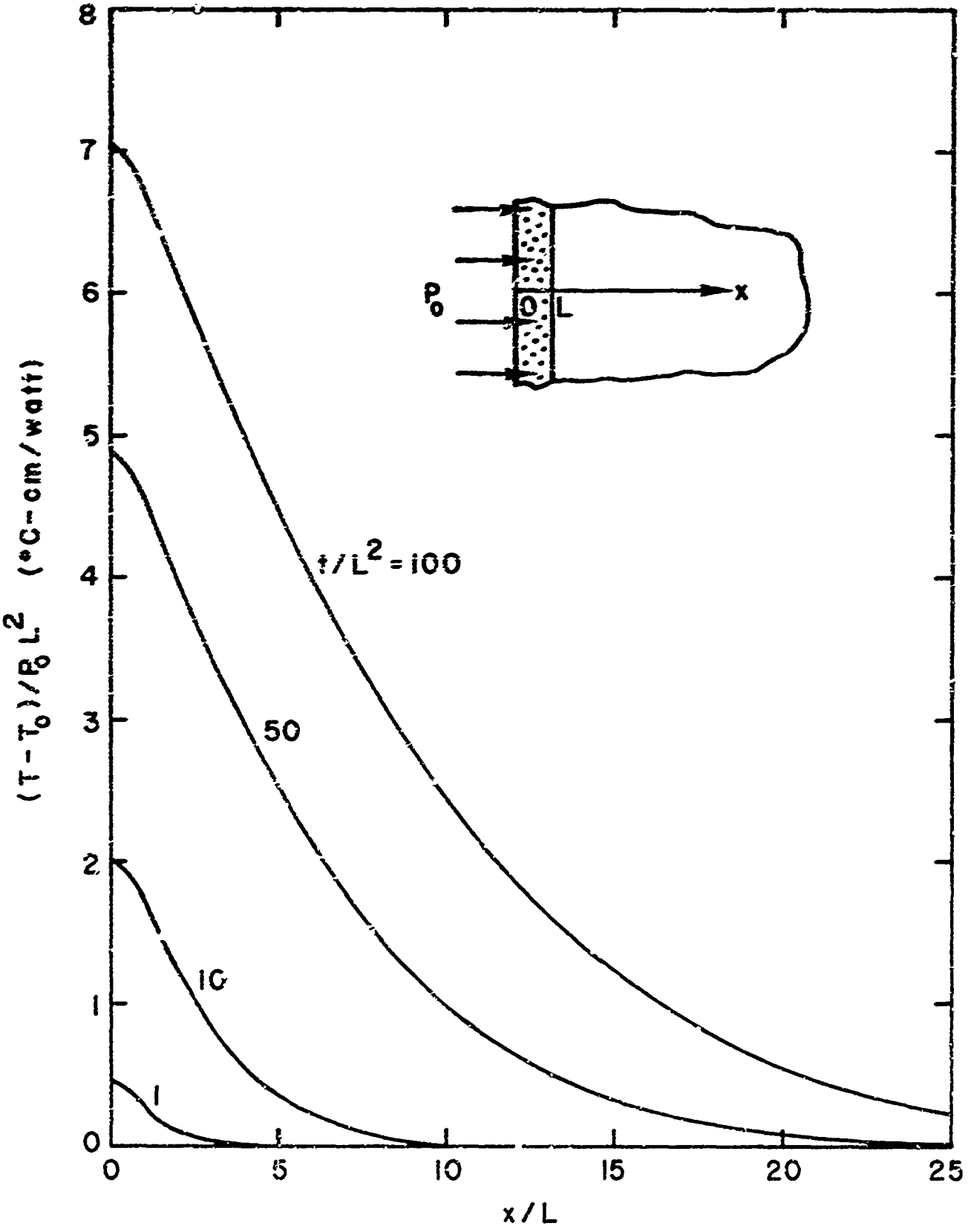


Fig. 5. Temperature distribution in a semi-infinite solid due to a power density P_o watts/cm^3 in a layer of thickness L at the surface.

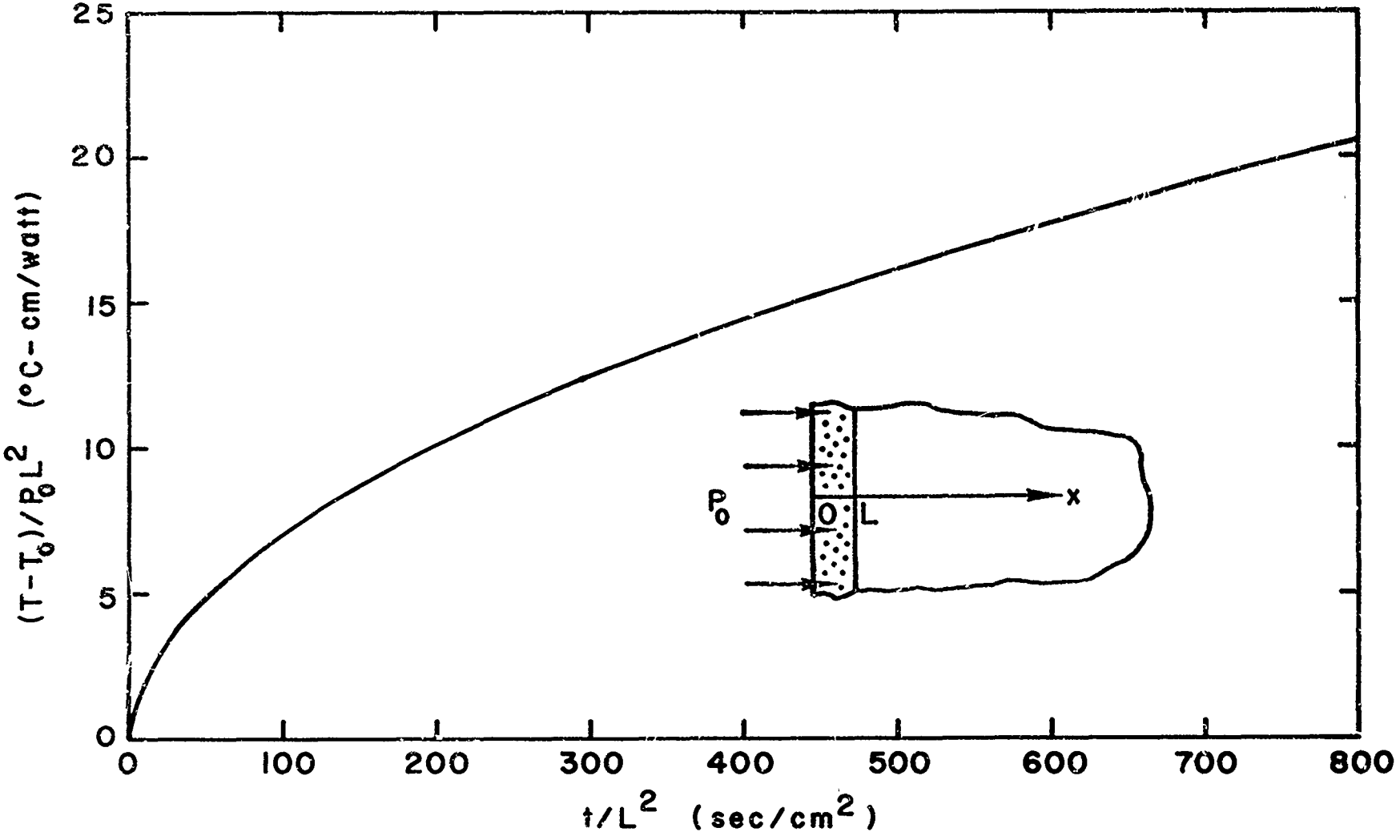
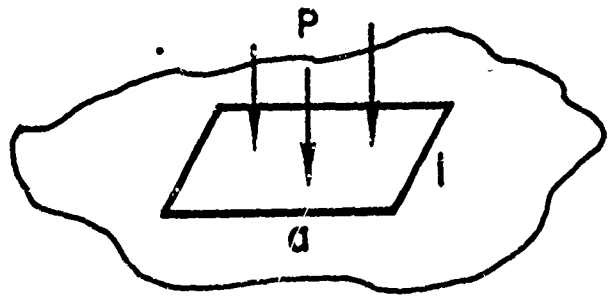


Fig. 6. Temperature rise at the surface of a semi-infinite solid with a power input P_o watts/cm^3 in a layer of thickness L at the surface.

case the ordinate in the two figures x40 gives the temperature rise, and the parameter in Figure 5 is read directly in μsec.

These figures agree with the more general results of Mosekilde, et.al.,[3] who have solved the same problem but for a semiconductor of finite thickness with an ideal heat sink.

5. Power into a rectangular area on the surface of a semi-infinite solid.



This represents the case of power dissipation by an active element in a thin layer at the surface of a chip. The assumptions again are that the edges of the die are far removed from the active element and that the die-header interface is far enough away during times of interest.

The solution takes into account the flow of heat laterally as well as vertically down into the silicon. The mathematics becomes complicated, but a simple expression can be found for the maximum temperature rise, T_m, which occurs at the surface at the center of the active element. This expression is

$$T_m - T_o = \frac{P}{\pi k a \ell} \{a \sinh^{-1} \frac{\ell}{a} + \ell \sinh^{-1} \frac{a}{\ell}\} \qquad (8)$$

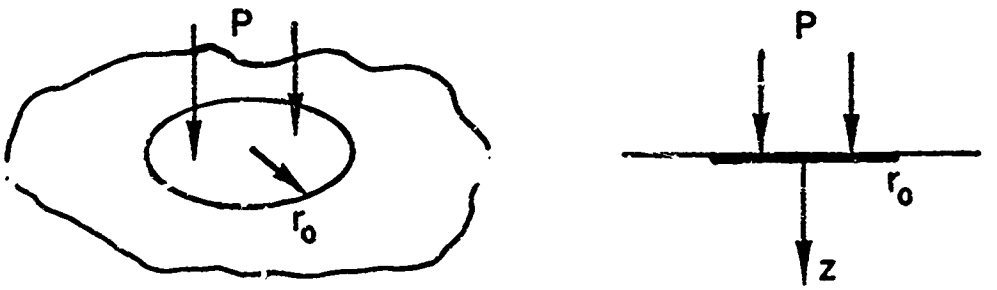
For the case $a = \ell$

$$T_m - T_o = 3.87 \times 10^3 \frac{P(\text{watts})}{a(\mu m)} \qquad (9)$$

Steady state temperature distributions in a finite chip with the

heat source at the center of the top surface and an ideal heat sink at the bottom surface have been calculated by Lindsted and Murty.[4] Normalized curves have been plotted for the maximum temperature rise for a range of geometries.

6. Power into a circular area on the surface of a semi-infinite solid.



This is similar to case 5; however, for this geometry it is relatively simple to determine T(z,t) along the axis of symmetry r = 0. This not only gives the maximum temperature rise, but allows one to examine the temperature distribution as a function of time and make a judgement as to when die thickness and edge effects are no longer negligible.

The solution is

$$T-T_o = \frac{2\sqrt{Dt}\,P}{\pi k r_o^2}\left\{\text{ierfc}\,\frac{z}{2\sqrt{Dt}} - \text{ierfc}\,\frac{(z^2+r_o^2)^{\frac{1}{2}}}{2\sqrt{Dt}}\right\} \qquad (10)$$

where $\text{ierfc}\,\mu = \frac{1}{\sqrt{\pi}} e^{-\mu^2} - \mu\,\text{erfc}\,\mu$.

This equation has several assymptotic approximations. When z = 0, T = T_m and

$$T_m - T_o \approx \frac{0.235\sqrt{t}\,P_o}{r_o^2} \text{ for } \frac{\sqrt{Dt}}{r_o} < 0.4 \qquad (11)$$

In this equation and those which follow t is in seconds, P is in watts, and spatial dimensions are in centimeters. This is the same as the solution for

case 2 if x approaches zero in Equation (5). This means that in a silicon chip where D = 0.9 cm^2/sec the lateral heat flow away from the edge of the circular region has no effect on T_m if $t<0.2r_o^2$.

In the steady state

$$T_m - T_o = 0.22 \frac{P}{r_o}, \qquad (12)$$

valid for $\sqrt{Dt}/r_o > 5.0$.

For large z and t

$$T - T_o \approx 0.11 \frac{P}{z} \quad \text{for} \quad \frac{\sqrt{Dt}}{r_o} > 50. \quad \text{and } 3/r_o > 5.0 \qquad (13)$$

Equation (10) is plotted in normalized form in Figure 7 and for z = 0 in Figure 8. If P = 1 watt and r_o = 1 mil, the ordinate in Figure 7 multiplied by 400 gives the temperature rise in °C. The curve with parameter 1 is the temperature distribution at t = 6.25 μsec while the curve labelled 50 corresponds to 16 msec, at which time the distribution is essentially steady-state. In the steady-state T_m is 88°C above ambient.

Steady-state solutions for a finite chip of cylindrical geometry are given by Kennedy.[5] Complete solutions for a chip of very large area but finite thickness mounted on a perfect heat sink are given in reference 6. If $r_o > 3L$, where L is the chip thickness, the solution is approximately that given by Equation (6), while if $L > 3r_o$ the solution is approximately Eq (12).

The solutions presented above are useful for estimating maximum temperatures and size effects when the temperature rise is not excessive. Unfortunately, the thermal properties of silicon, especially the conductivity, are functions of temperature. This raises serious questions about the

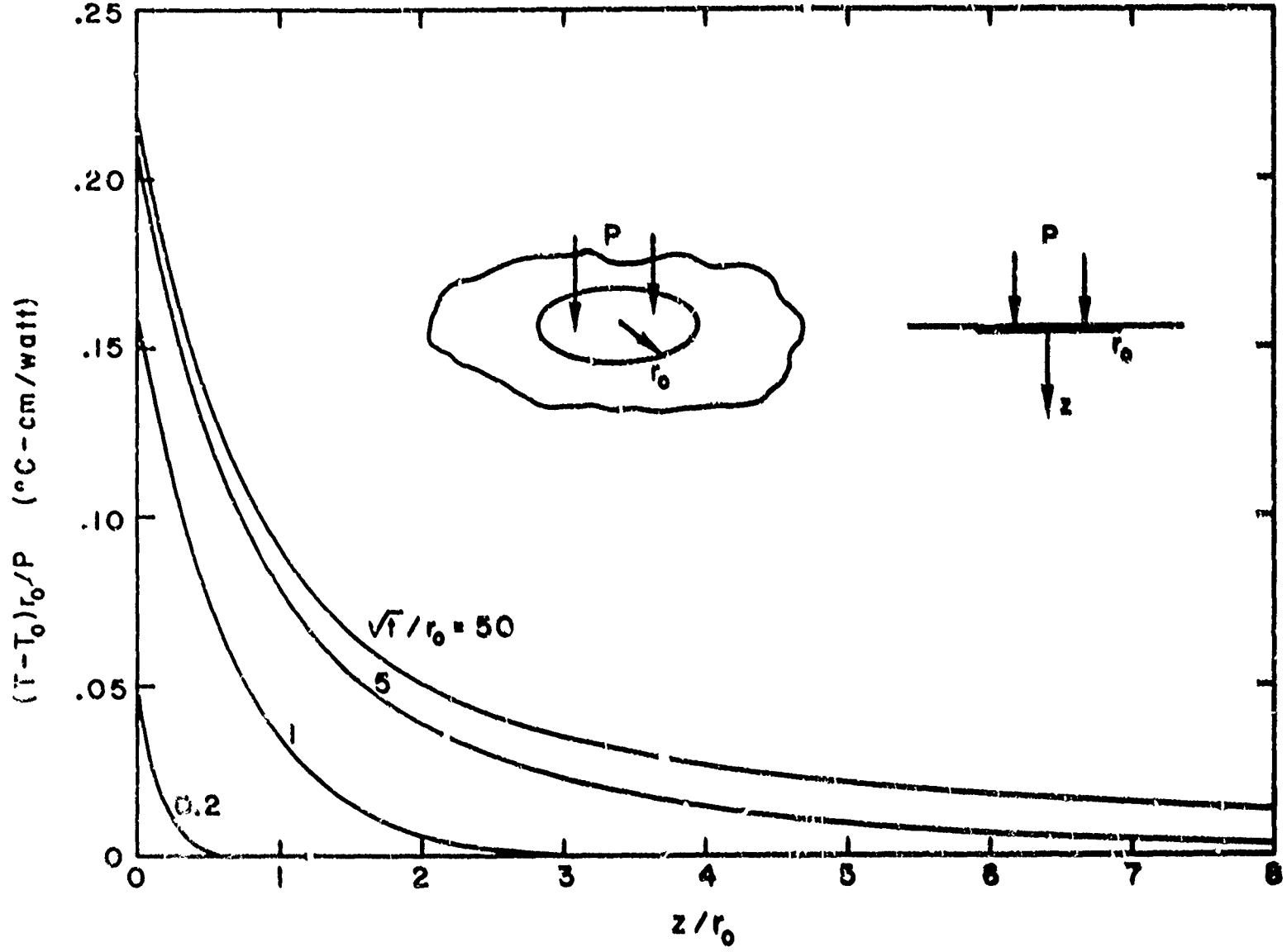


Fig. 7. Temperature distribution in a semi-infinite body due to a constant power applied at the surface over a circular area of radius r_0. The curve is the distribution along the perpendicular axis of the circular region.

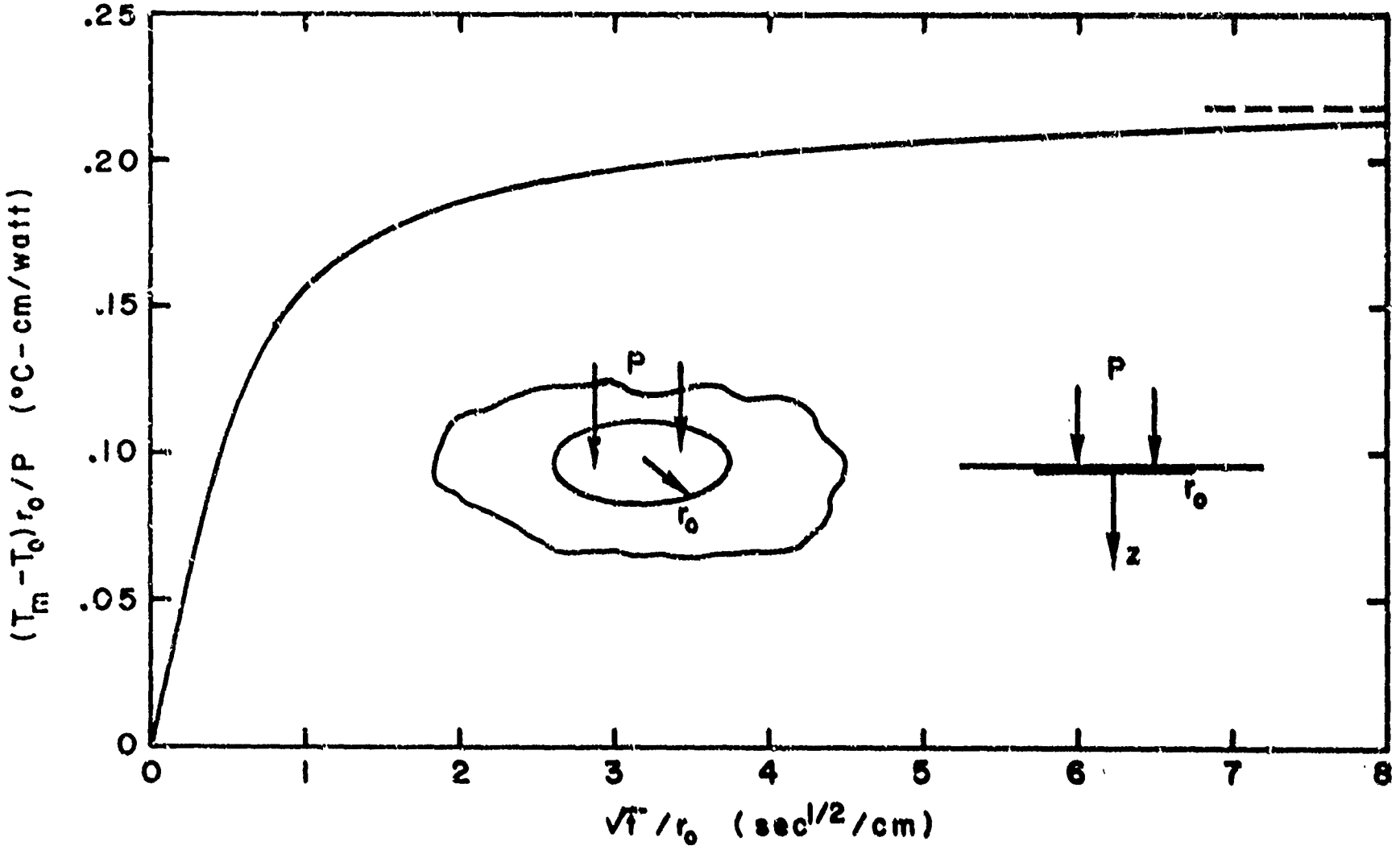


Fig. 8. Temperature rise in a semi-infinite solid due to power applied at the surface over a circular area of radius r_o. The temperature rise is the maximum value, at the center of the circular area. The dashed line at the right is the asymptotic value.

validity of using solutions to the linear heat equation. In the next subsection the variation of thermal conductivity with temperature will be discussed, and some of the difficulties associated with the use of an average value will be pointed out.

C. Thermal Conductivity of Silicon

The physics of the thermal conductivity of solids is well understood[7,8] and good qualitative and even quantitative agreement between theory and measurements is obtained for many materials. Any lack of quantitative agreement is due to mathematical complexities and not to any lack of understanding of the principles.

Thermal conduction in solids occurs by means of phonons, electrons, holes, electron-hole pairs, and photons. Lattice or phonon conduction is the only important mechanism in silicon at temperatures less than room temperature. The others become increasingly important at higher temperatures.

Finite phonon thermal conductivity is attained through the action of scattering mechanisms which keep the phonon distribution in thermal equilibrium with the lattice temperature. These mechanisms are: diffuse scattering from crystal boundaries, phonon-phonon scattering, isotope scattering, electron scattering, and scattering from crystal defects and imperfections. Scattering establishes a mean free path for phonon propagation which can be related to the thermal conductivity.

For an ideal Maxwell-Boltzmann gas the thermal conductivity is given by

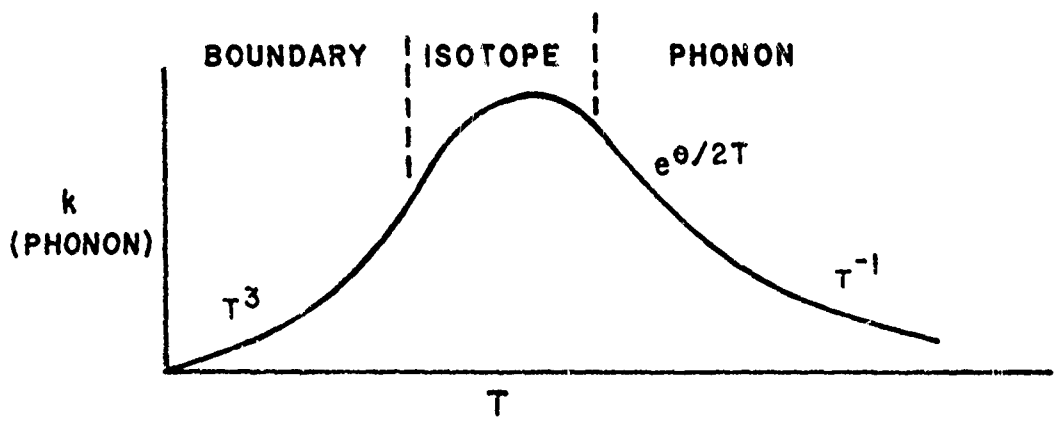
$$k = \frac{1}{3} cv\ell \tag{13}$$

where c is the specific heat at constant volume, v is the average velocity, and ℓ is the mean free path. How do the scattering mechanisms affect ℓ and

and hence k? At very low temperatures the number of phonons available to scatter other phonons is small. The electron (hole) population is small since there is little thermal generation and even impurity carriers may be frozen out. Isotope effects are of little importance because the phonons are of long wavelength. This leaves only boundaries and defects as effective scattering agents. In this case ℓ is fixed, and, since $c \propto T^3$ at low temperatures, $k \propto T$ and k approaches zero as T approaches zero.

As the temperature is increased, isotopes can be assumed to have increasing importance.* If we restrict ourselves to pure material for the moment, electron effects can still be neglected. Scattering by other phonons depends on the number of other phonons available, which depends on $e^{\theta/2T}$, where θ is the Debye temperature. At very high temperatures the phonon population varies as T. c is nearly constant at high temperatures, hence k varies as 1/T.

These interactions will cause the phonon portion of the thermal conductivity to vary as shown below.



Above room temperature other carriers besides phonons contribute to heat transport. Photon conduction is observed in pure, large bandgap semiconductors at high temperatures, for example GaAs above 800°K. It

* Natural silicon consists of about 92% atomic weight 28, 5% 29, and 3% 30.

actually should be treated as conduction by radiation, and previously this was shown to be small in silicon. Glassbrenner and Slack[9] were unable to detect a photon contribution up to 1580°K, within 100 degrees of the melting point.

Heat conduction by electrons (holes) occurs by the same process as heat conduction by electrons in metals. Contrary to metals, however, where nearly all of the heat is carried by the numerous free electrons, only about 6% of the thermal conductivity of silicon at the melting point is due to free carriers.

Electron-hole pair conduction is significant at high temperatures because of thermal generation of large densities of electron-hole pairs. These carriers diffuse (ambipolar diffusion) from regions of high temperature and concentration to regions of lower concentration. When recombination takes place energy roughly equal to the band gap energy is released. Glassbrenner and Slack have found ambipolar diffusion to comprise 32% of the thermal conductivity at the melting point.

The thermal conductivity of pure silicon is plotted in Figure 9. The values are taken from Glassbrenner and Slack. The variation is as expected, and indeed Glassbrenner and Slack claim good agreement with theoretical calculations. The T^3 behavior at low temperatures is suppressed by the scale on the abscissa, chosen to better display the conductiviy at higher temperatures.

The effects of doping with boron or phosphorus have been measured by Slack.[10] He reports that the room temperature conductivity is reduced from 1.56 to 1.2 watts/cm-°K for phosphorus doping of $2 \times 10^{19}/cm^3$. Below room temperature the relative change is even greater. At the heaviest

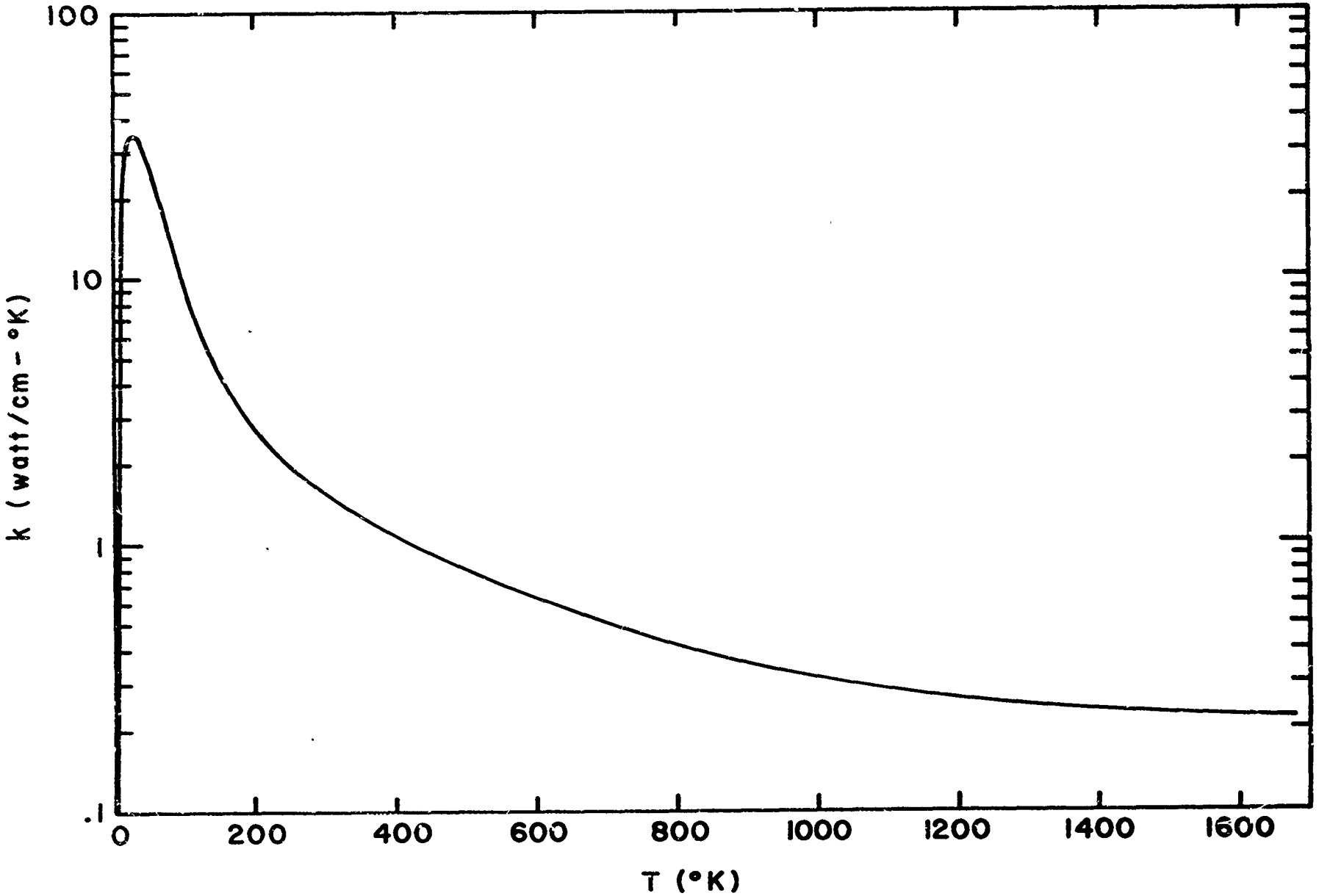


Fig. 9. Thermal conductivity of pure silicon as a function of absolute temperature.

doping, boron at 3×10^{20}, k at room temperature is reduced to about 0.47 watts/cm-°K. He also reports on earlier work by other workers of boron doping at $4 \times 10^{16}/cm^3$. In this case, k was reduced only at temperatures less than 80°K.

The extreme variation of k in Figure 9 causes justified reluctance in applying solutions to the linear heat flow equation even if an average value is assumed. It has been common practice to use about 0.9 watts/cm-°K. It is worthwhile to examine carefully a typical case to see the difficulties which can be encountered.

In one such case an estimate of the maximum temperature rise under certain conditions was needed. The power input was 2.8 watts into a thin layer at the surface with area 20 μm x 40 μm. The model in example 5 applies. Using Equation (8) the first estimate of the steady-state temperature rise is 343°C. If we take T_o = 300°K and use an average value of k, that is

$$k = \tfrac{1}{2}(k(300)+k(T_m)) \tag{14}$$

trial and error yields T_m = 500°C. If instead we use k at the average temperature,

$$k = k[\tfrac{1}{2}(300+T_m)] \tag{15}$$

then T_m = 1050°C. A graphical solution is shown in Figure 10. Equation (14) is plotted as the upper dashed line. It is the average of k(300) = 1.56 and $k(T_m)$ which is simply Glassbrenner and Slack's data of Figure 9 replotted. The vertical lines with arrowheads indicate the two curves which are averaged. The intersection of the dotted line with the solid curve $k = 498/T_m$, which is a plot of Equation (8), gives the solution T_m = 497°C.

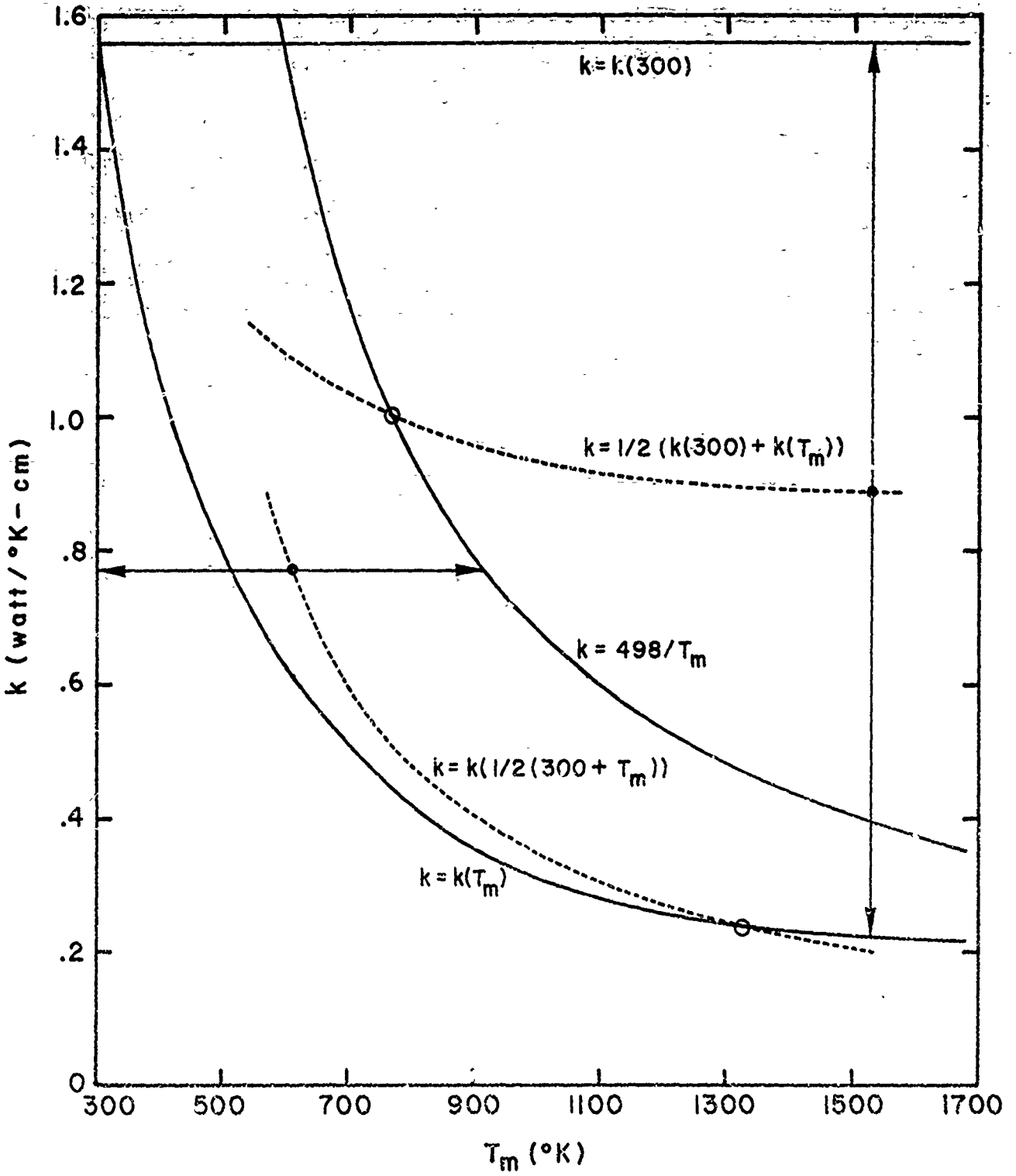


Fig. 10. A graphical solution for the temperature rise in an integrated circuit with two different approximations for the thermal conductivity. See text for explanation.

Equation (15) states that k should be found at a temperature which is the average of 300°K and T_m. This average is the lower dotted line in Figure 10. The intersection with $k = k(T_m)$ gives this solution as 1057°C.

The disparity between the two solutions is a serious impediment to the use of solutions to the linearized equation in general cases. Numerical solutions must be resorted to even though these are more difficult, subject to inaccuracies, and not amenable to the usual normalization. The next section is devoted to a discussion of numerical methods and the presentation of some results of a computer program for solving heat flow problems.

D. Numerical Solutions to Heat Flow Problems

A simple approach to the solution of the heat equation is to quantize the space and time variables and write a heat balance equation for each small volume. Knowing the net heat flow into an incremental volume enables one to calculate the temperature rise during the next incremental time interval. If the space and time increments are small enough an accurate solution to the differential equation can be obtained. The thermal properties of the material can be adjusted as the temperature changes, and if the temperature does not change greatly during one time step good approximations to the physical problem can be obtained.

It is sometimes helpful to visualize heat flow by its analogy with an electrical circuit. The analagous quantities are:

thermal resistance	-	electrical resistance
thermal heat capacity	-	electrical capacitance
heat flow	-	current flow
temperature	-	voltage

For simplicity a one-dimensional example will be illustrated, with

the space variable x divided into equal regions. We will speak of a node as the point at the center of each region. The node temperature is measured in degrees Kelvin or Celsius. The temperature difference between adjacent nodes will result in a flow of heat energy, determined by the thermal resistance between the nodes. The mass and specific heat of each region determine how much the temperature will change as a result of net heat flow during the next time interval.

The electrical analog to the thermal model is shown in Figure 11. The node temperatures are analagous to the node voltages in the electrical circuit. The circuit capacitors are connected between the nodes and ground. The reason for this is that ground potential represents the ambient temperature. In the same way that we are really interested only in the temperature rise above ambient, or the temperature difference between nodes, so only the node voltages with respect to some arbitrary reference or the difference in potential between nodes is of importance. Note that the capacitors are not connected between nodes. Power dissipation in a region caused by i^2R heating, heating of a depletion layer, etc., is represented by a current source directed into the node.

Writing an expression for the heat flow into the nth node in the thermal model we obtain

$$\frac{kA(T_{n-1}-T_n)}{\Delta x} - \frac{kA(T_n-T_{n+1})}{\Delta x} + qA\Delta x = 0$$

where q is the heat density in watts/cm^3 being supplied to the nth node. Divide this equation by the specific heat mass of the nth region and in a time Δt the temperature rise will be

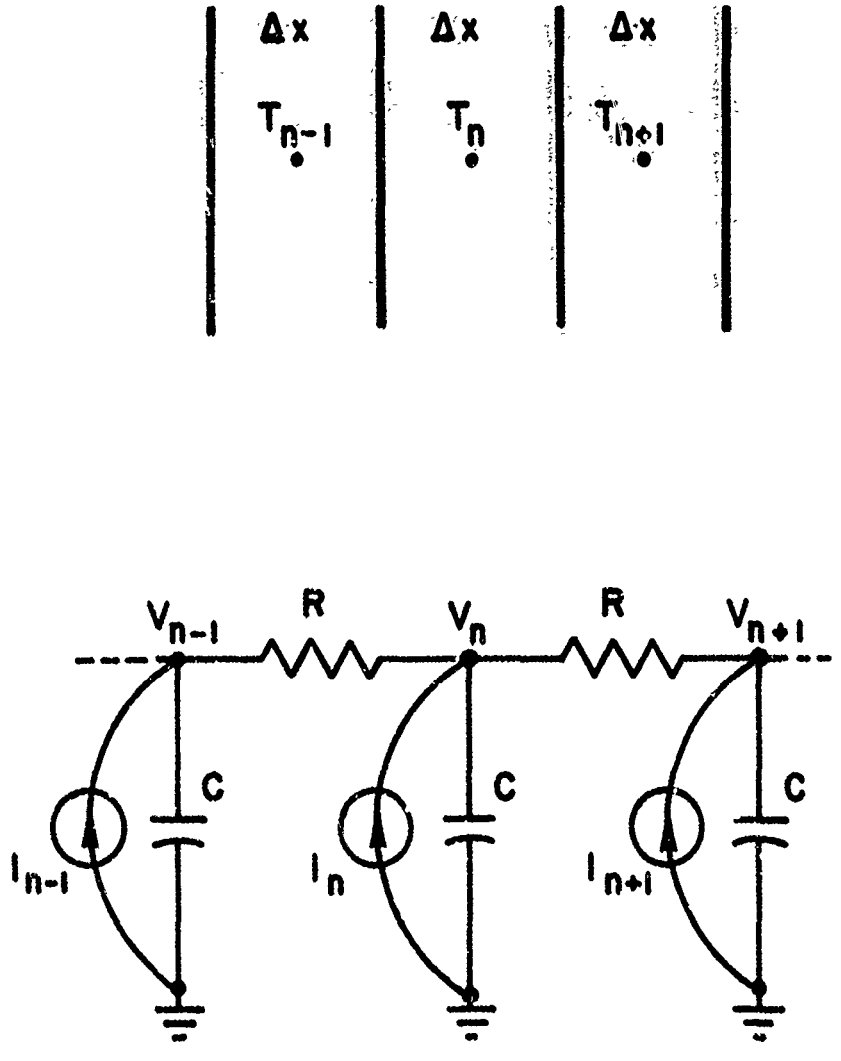


Fig. 11. Analogy between the thermal model (top) and the electrical model (bottom).

$$\Delta T = \frac{\Delta t}{\rho c A \Delta x}\left[\frac{kA}{\Delta x}\left(T_{n-1}-2T_n+T_{n+1}\right)+qA\Delta x\right] \qquad (16)$$

Since the initial conditions for the problem are known, Equation (16) can be immediately solved for each node. This gives the new temperature distribution at Δt. This technique is known as the "explicit" formulation of the finite difference equation for heat flow, or the Bender-Schmidt method. Equation (16) is derived in a more formal way in texts on finite difference equations.

For problems where the heat source is at one end of the structure, say at node 1, convergence is faster if the temperature rise of the nth node is calculated in terms of the new temperature of node n-1. As an example of this approach consider a diode with a cross section area of $10^{-4}cm^2$ dissipating 50 watts in a 2 μm layer at the surface. The problem was solved in one dimension only, with the variation in thermal conductivity approximated by a simple analytic expression. The temperature distribution as a function of time is shown in Figure 12. For comparison the dotted curve shows the temperature distribution with k constant and equal to the room temperature value. The difference is appreciable and once again emphasizes the need for realistic solutions when large temperature rises are encountered.

This method has been used to investigate the temperature rise in aluminum metallization runs on integrated circuits with good results.[11,12] The technique is simple, yet very flexible. Not only can the conductivity and specific heat be functions of temperature, but the power sources can be functions of temperature and time. There is considerable freedom in choosing node sizes and different materials can be accommodated. The algorithm is fast. It does suffer from one drawback; however, the solution is unstable

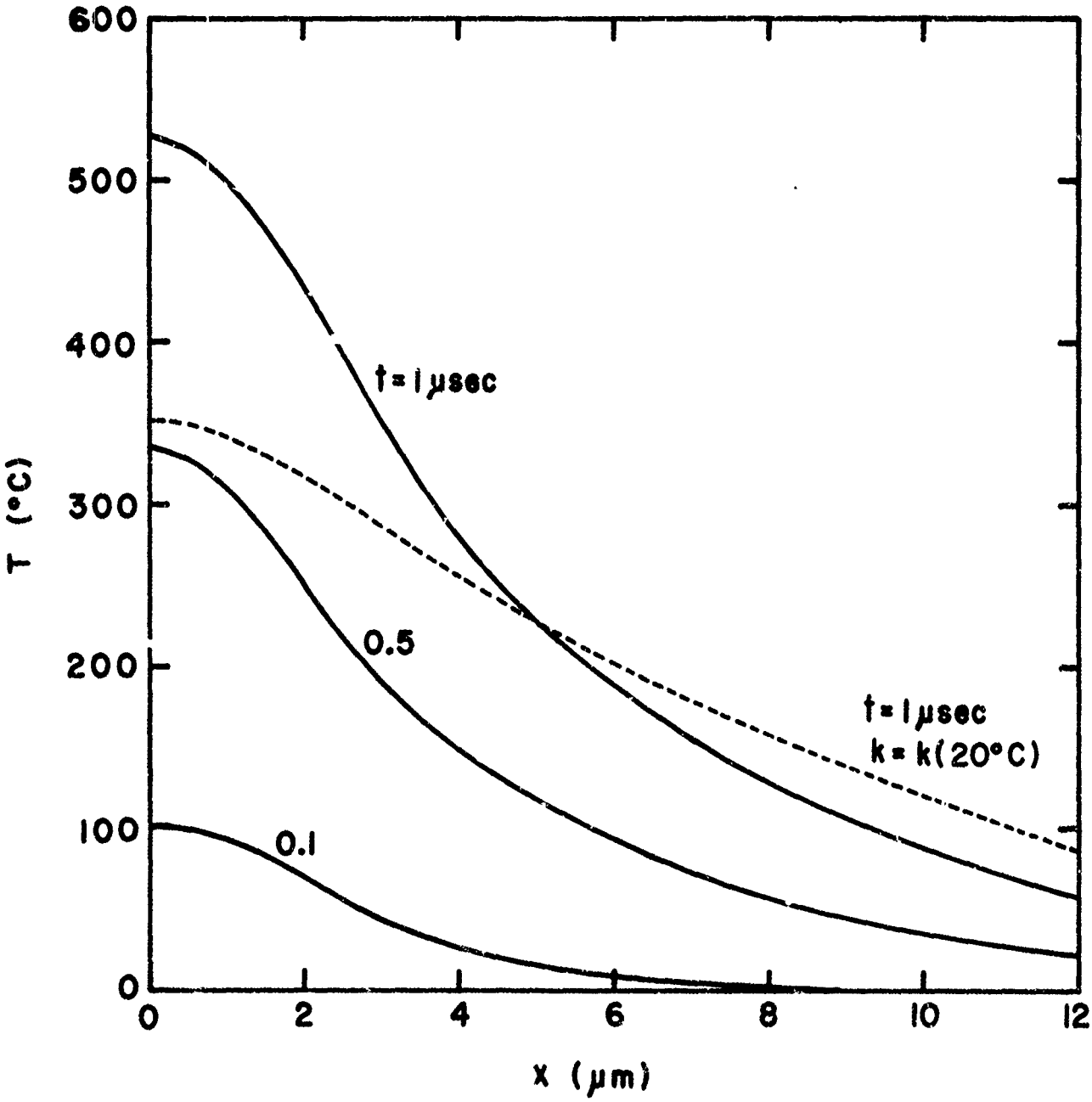


Fig. 12. Temperature distribution in a silicon diode, solved by an explicit, finite difference method. The dotted line is the solution at 1 μsec, assuming a constant value for the thermal conductivity.

if the time increment is too large. Once the node sizes have been selected Δt must satisfy the condition

$$\Delta t \leq \frac{(\Delta x)^2 \rho c}{2k} \tag{17}$$

If thermal oxide or metallization runs are to be included in the problem Δx is of the order of tenths of microns. Then Δt must be smaller than tenths of nanoseconds. If the final solution is needed for times of the order of milliseconds not only are enormous computing times required, but the solution becomes suspect due to numerical inaccuracies.

Thus in certain situations the stability condition proves to be such a limitation that other, less intuitive, but more powerful methods are desirable. Before discussing one such method we will digress for a moment to discuss general techniques for numerical solutions to the heat equation.

The finite difference method is the oldest and most widely used approach to the solution of heat conduction problems. Monte Carlo methods are occassionally used, and the finite element method is a powerful way of handling a certain class of problems. Neverthelss, finite differences have many desirable features and are usually used. Many excellent texts are available on the subject.[13,14]

The heat diffusion equation is one type of parabolic differential equation. The finite difference approach aims to approximate the first order derivative and the second order derivative by means of Taylor series expansions and obtain algebraic equations. Rewriting the heat equation in its linearized one-dimensional form

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial x^2} + \frac{q}{\rho c} \tag{18}$$

$T(x,t)$ can be expanded in the <u>forward</u> direction

$$T(x+\Delta x,t) = T(x,t) + \Delta x \frac{\partial T}{\partial x} + \frac{(\Delta x)^2}{2!} \frac{\partial^2 T}{\partial x^2} + \ldots . \qquad (19)$$

or in the backward direction

$$T(x-\Delta x,t) = T(x,t) - \Delta x \frac{\partial T}{\partial x} + \frac{(\Delta x)^2}{2!} \frac{\partial^2 T}{\partial x^2} - \ldots . \qquad (20)$$

If we neglect higher order terms, Equations (19) and (20) can be solved to yield the central difference approximation to the second derivative

$$\frac{\partial^2 T}{\partial x^2} \approx \frac{1}{(\Delta x)^2} [T(x+\Delta x,t) - 2T(x,t) + T(x-\Delta x,t)] . \qquad (21)$$

The forward difference approximation of the time derivative is

$$\frac{\partial T}{\partial t} = \frac{1}{\Delta t} [T(x,t+\Delta t) - T(x,t)] . \qquad (22)$$

Equations (21) and (22) lead directly to the explicit heat flow equation, namely Equation (16).

The backward difference approximation of the time derivative is

$$\frac{\partial T}{\partial t} = \frac{1}{\Delta t} [T(x,t) - T(x,t-\Delta t)] . \qquad (23)$$

This leads to a somewhat different finite difference formulation of Equation (18), namely

$$T_n(t+\Delta t) - T_n(t) = \frac{\Delta t}{\rho c A \Delta x} \left\{ \frac{kA}{\Delta x} [T_{n+1}(t+\Delta t) - 2T_n(t+\Delta t) + T_{n-1}(t+\Delta t)] + qA\Delta x \right\} . \qquad (24)$$

The difference is that the rise in temperature at a node is written in terms of the new node temperatures, as yet unknown. We are lead to write a group of n simultaneous equations. This is called the implicit method or Liebman's method. There seems to be no particular advantage to this procedure until we realize that there is no stability problem. The method is inherently stable for any time step size. There are still questions of accuracy to be

investigated, but not of stability. Using this approach one can avoid stability problems and choose a value of Δt based on other considerations such as accuracy.

A number of variations of the previous development exist for the purpose of speeding convergence or increasing accuracy. These typically consist of a weighted average of forward and backward differences. See reference 14 for further details.

Convergence and accuracy are problems of which little of a quantitative nature can be said. In general, if Δt and Δx are made sufficiently small the numerical solution will approach the true solution. Trial and error must often be resorted to. Having solved a problem the solution can be compared to that of the same problem by halving Δt and Δx. If the change in the solution is small one is usually justified in assuming that the error is small.

The main drawback of the implicit method is the need to invert the matrix - repeatedly if variable coefficients are to be allowed. Fortunately the matrix is sparse. Gauss-Siedel or some similar iterative method of solution is usually recommended to solve the simultaneous equations. Unfortunately, round-off error can sometimes be a problem. Since the matrix has many null terms it is worthwhile to consider other, more direct methods than iterative schemes.

Richtmeyer[14] describes a very fast way of solving the set of Equations (24). Each equation has at the most three unknown temperatures, and can be written in the form

$$A_{jj-1}T_{j-1} + A_{jj}T_j + A_{jj+1}T_{j+1} = C_j \qquad (25)$$

If we write

$$T_j = E_j T_{j+1} + F_j \tag{26}$$

Then by substitution into (25)

$$E_j = - \frac{A_{jj+1}}{A_{jj} + A_{jj-1} E_{j-1}} \tag{27a}$$

$$F_j = \frac{C_j - A_{jj-1} F_{j-1}}{A_{jj} + A_{jj-1} E_{j-1}} \tag{27a}$$

When $j = 1$, Equation (25) has only 2 unknown temperatures T_1 and T_2 and

$$T_1 = \frac{C_1}{A_{11}} - \frac{A_{12}}{A_{11}} T_2$$

$$\text{so that} \qquad F_1 = \frac{C_1}{A_{11}} \qquad \text{and} \qquad E_1 = - \frac{A_{12}}{A_{11}} \tag{28}$$

Now all values of E_j and F_j can be calculated from (27).

If we assume that the nth node remains at $T_n = T_o$, then the last equation of (25) can be solved for T_{n-1}.

$$T_{n-1} = E_{n-1} T_o + F_{n-1}$$

Now Equation (26) will give all the unknown node temperatures.

This algorithm was found to be much more accurate than the Gauss-Seidel method for coefficients encountered in a realistic problem. In addition it was faster by a factor of twenty times in certain problems.

A computer program was written to solve one-dimensional problems in heat conduction. A program listing appears in Appendix I. The program is called THAP - for Transient Heat Analysis Program. It makes use of the implicit method and Richtmeyer's algorithm. As presently written the

program will solve problems with up to 20 nodes of arbitrary sizes consisting of silicon, silicon dioxide, and aluminum in any order. The initial temperature must be uniform, but can have any value. The thermal conductivity and specific heat of all three materials are functions of temperature.

Examples of problems solved with THAP are shown in Figures 13 and 14. Figure 13 shows the temperature rise in a silicon chip when power is supplied uniformly throughout a surface layer 4 microns thick. The power density, 1.5×10^5 watts/cm^2-μm, corresponds to 50 watts over an area about 90 microns by 90 microns. The peak temperature increases indefinitely in this example, in fact reaching the melting point of silicon in about 5 μsec. In 1 μsec the temperature at x = 10 μm is about 25% of T_m. This should not be taken to mean, however, that some other boundary condition at x = 10 μm, such as a plated heat sink, would have a pronounced effect on T_m. It would have some effect, of course, but quantitative answers can only be obtained by running the problem again.

One other precaution should be pointed out when using the results of computer calculations. In Figure 13 the axes are not normalized. Since the problem is non-linear the usual normalized parameters cannot be used, and caution must be exercised when attempting to extrapolate the curves to fit new situations. If, for example, the temperature rise is 400°C above ambient in time t_o, the rise is not doubled if the input power is doubled. In fact the temperature rise will be more than doubled in a typical case.

The same problem has been rerun in Figure 14, except that the surface has been covered by 0.5 microns of silicon dioxide with 1 micron of aluminum on top of that. It is evident that the added thermal capacitance of the surface layers has lowered T_m appreciably, except for times less than

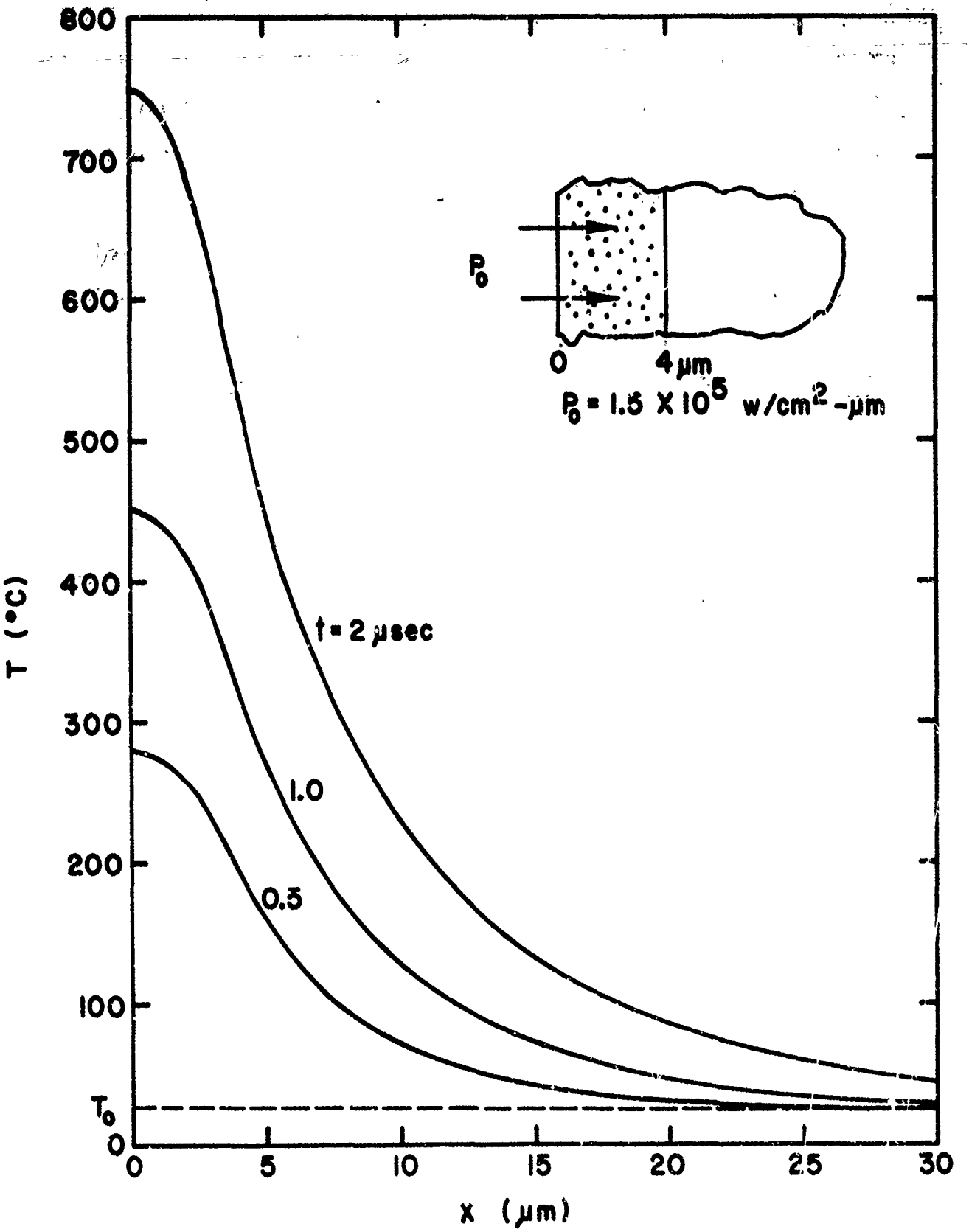


Fig. 13. Temperature distribution in silicon due to power applied in a 4 μm layer at the surface. Solution is by the implicit finite difference method using Richtmyer's algorithm and temperature dependent thermal properties.

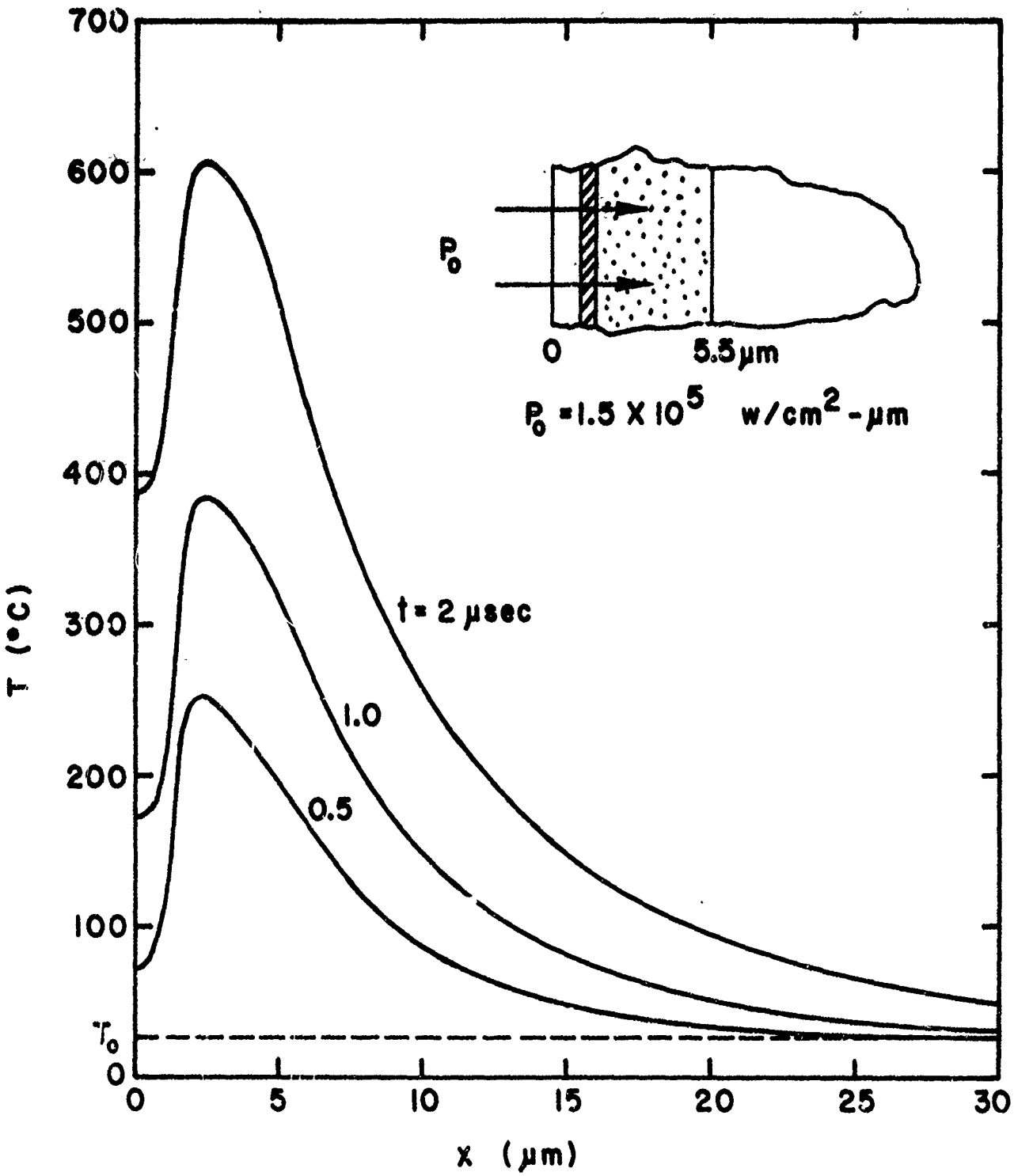


Fig. 14. Temperature distribution under the same conditions as Fig. 13, but with a 0.5 μm layer of silicon dioxide and a 1 μm layer of aluminum at the surface.

0.5 μsec. In the passive silicon, x>5.5 μm, the temperature is nearly unaffected by the added layers, at least for the times shown.

A third example, not plotted here, was exactly the same as the problem in Figure 14, except that the power was dissipated in a 1 micron layer between 3 and 4 microns from the surface, simulating power dissipation in a depletion layer beneath the surface. The temperature distribution was somewhat different, of course, but T_m was almost identical. In other words, fine detail in the power distribution was not important. A more quantitative estimate of the importance of detail in the model is obtained by calculating the diffusion length $L_D = (Dt)^{\frac{1}{2}}$, where D is the diffusion constant and t is the time. For silicon D is about 1 cm^2/sec. If t = 1 μsec, L_D = 10 μm. For a time period of 1 μsec heat spreads through distances of about 10 μm. This was observed in the case of Figure 13. It explains why the exact distribution of power within a layer less than 10 μm thick is not important for t of the order of 1 μsec.

Further details of the effect of single surface layers on peak temperature are given in Figure 15. Here again we have a one-dimensional problem, with a power density of 1.5×10^5 watts/cm^2-μm in a 4 μm layer. The peak temperature is plotted as a function of time. The dotted lines correspond to different thicknesses of SiO_2 on the surface. The line labelled 5 μm SiO_2 also holds for any thickness greater than 5 μm in the time scale given. This implies that glassivation layers 5 μm or thicker insulate the silicon from any possible heat-sinking effects of metal deposited on top of glass (for pulses less than about 10 μsec).

The solid lines, terminated at 660°C show the effects of aluminum deposited directly on the silicon. The decrease in T_m is quite pronounced.

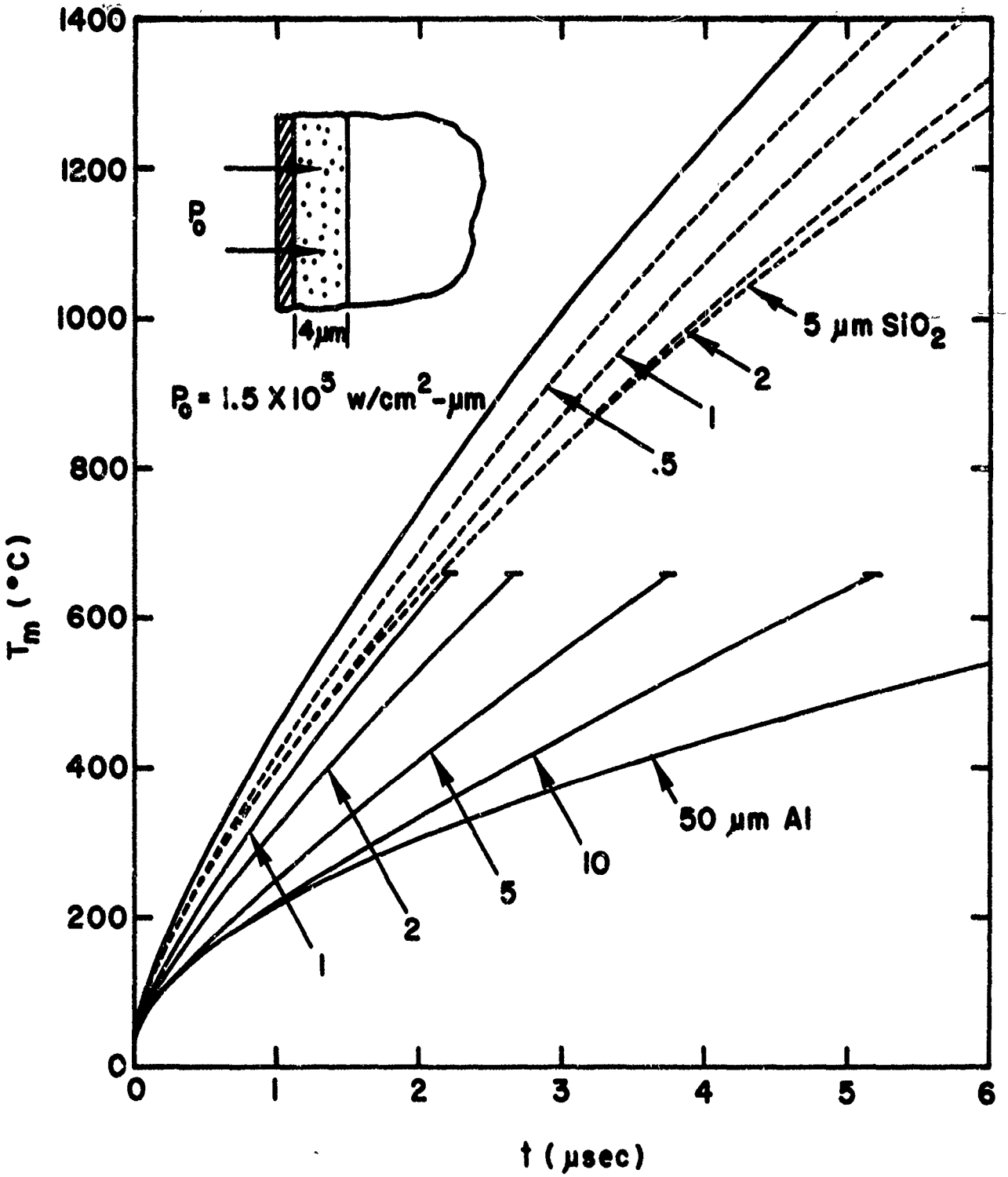


Fig. 15. Effect of surface layers on peak temperature rise in silicon. Power is applied in a 4 μm thickness, and the layers are either silicon dioxide or aluminum.

For times less than 6 μsec the 50 μm layer is the same as an infinite layer and gives a rough idea of the effect of a plated heat sink or a Schottky barrier on T_m. Since copper and gold have greater heat conductivity than aluminum T_m would be even lower than given in the figure.

More detailed plots near the origin show that the SiO_2 layer has a negligible effect on T_m when t is less than 100 nsec, and the same is true for aluminum when t is less than 20 nsec.

When both aluminum and SiO_2 are present, the results are more complicated. Figure 16 shows the peak temperature after 1.5 μsec as a function of the oxide thickness with the aluminum thickness as a parameter. As expected, the effect of the aluminum becomes less and less as the oxide thickness is increased. A 1 μm thickness of SiO_2 is nearly the same as an infinite layer, as can be verified in Figure 15. Of special interest is the fact that a 1 μm layer of aluminum lowers T_m to about 500°C irrespective of the oxide thickness. This rather surprising result means that a typical metallization run has about the same heat sinking capability whether it is over emitter oxide or over very thick oxide! Again it should be pointed out that this observation is only valid for one-dimensional heat flow and the other particular circumstances of Figure 16.

In Figures 17-19 the power density to reach a certain T_m as a function of time is plotted. The curves are probably more useful in inverted form, that is, given a power density, how long does it take to reach a certain temperature. The three figures are for power inputs in a 2, 3, and 4 μm layer, respectively. This is the range of base diffusions in most of the circuits tested experimentally. A layer of silicon dioxide 0.4 μm thick covers the surface, this being the average between base and emitter oxide

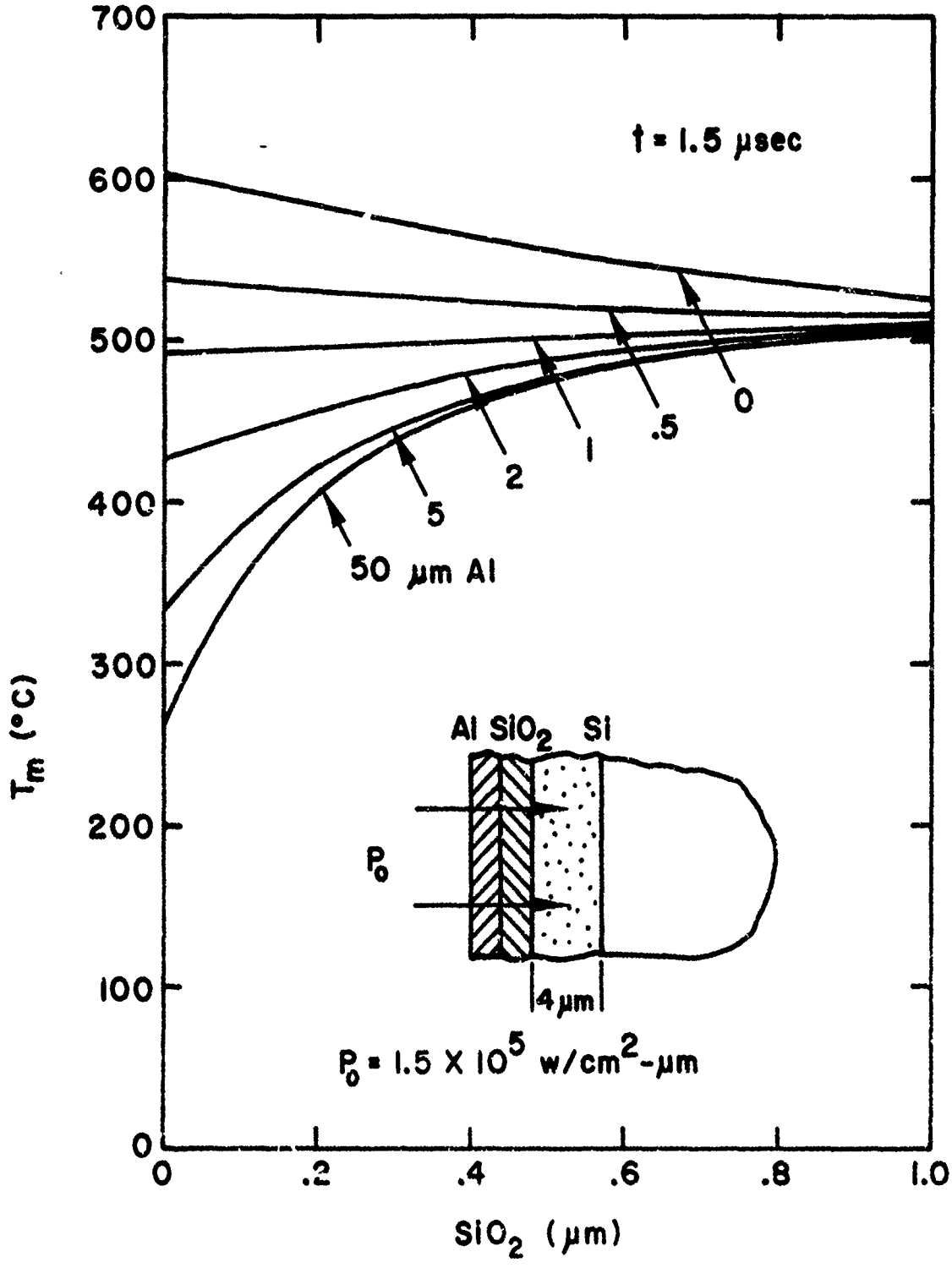


Fig. 16. Peak temperature rise in silicon after 1.5 μsec with variable thicknesses of silicon dioxide and aluminum.

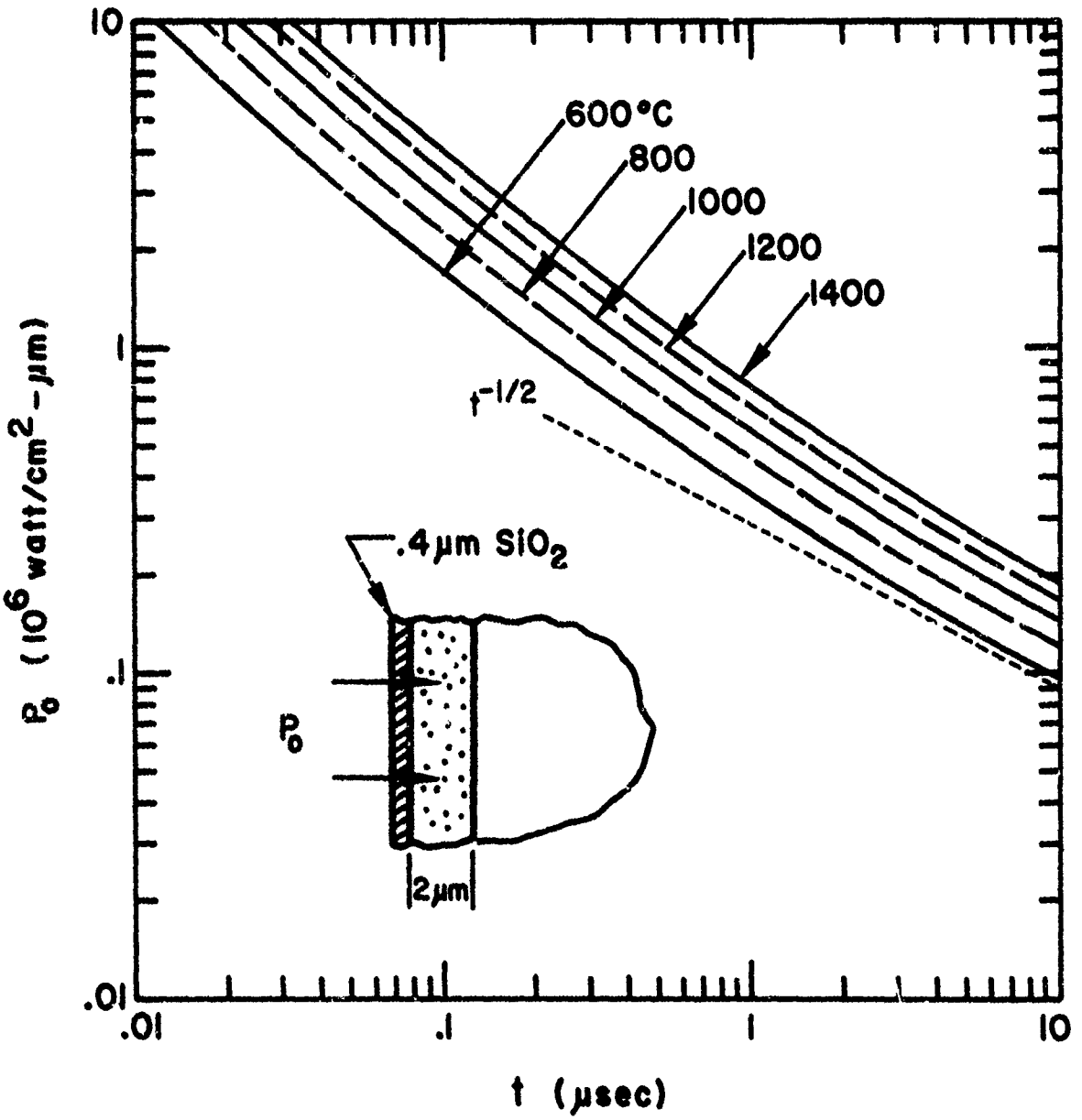


Fig. 17. Power versus time curves to reach the maximum temperature indicated. Power is applied to a 2 μm layer, with a 0.4 μm layer of silicon dioxide on the surface.

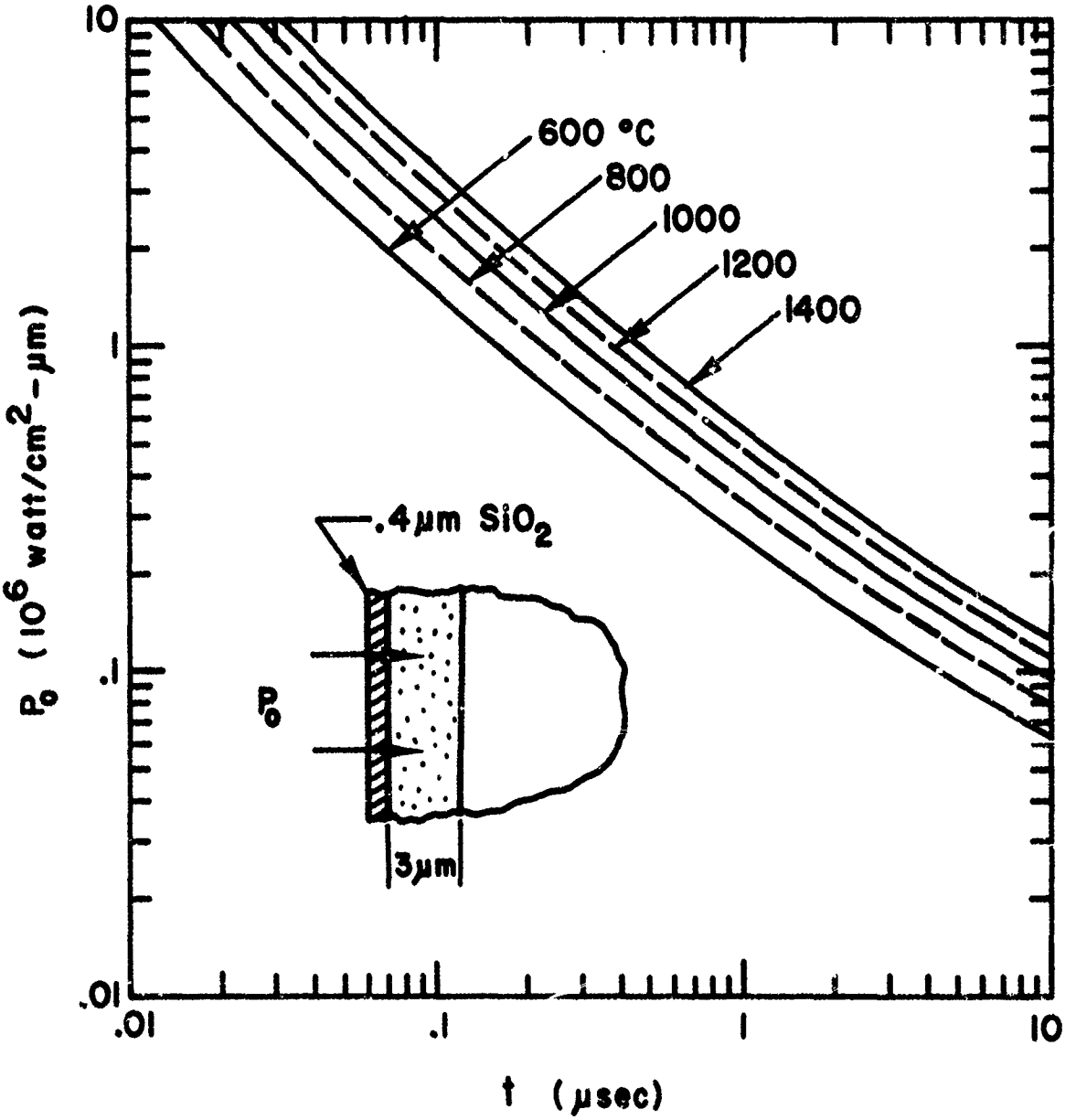


Fig. 18. Same as Figure 17, with the power applied to a 3 μm layer.

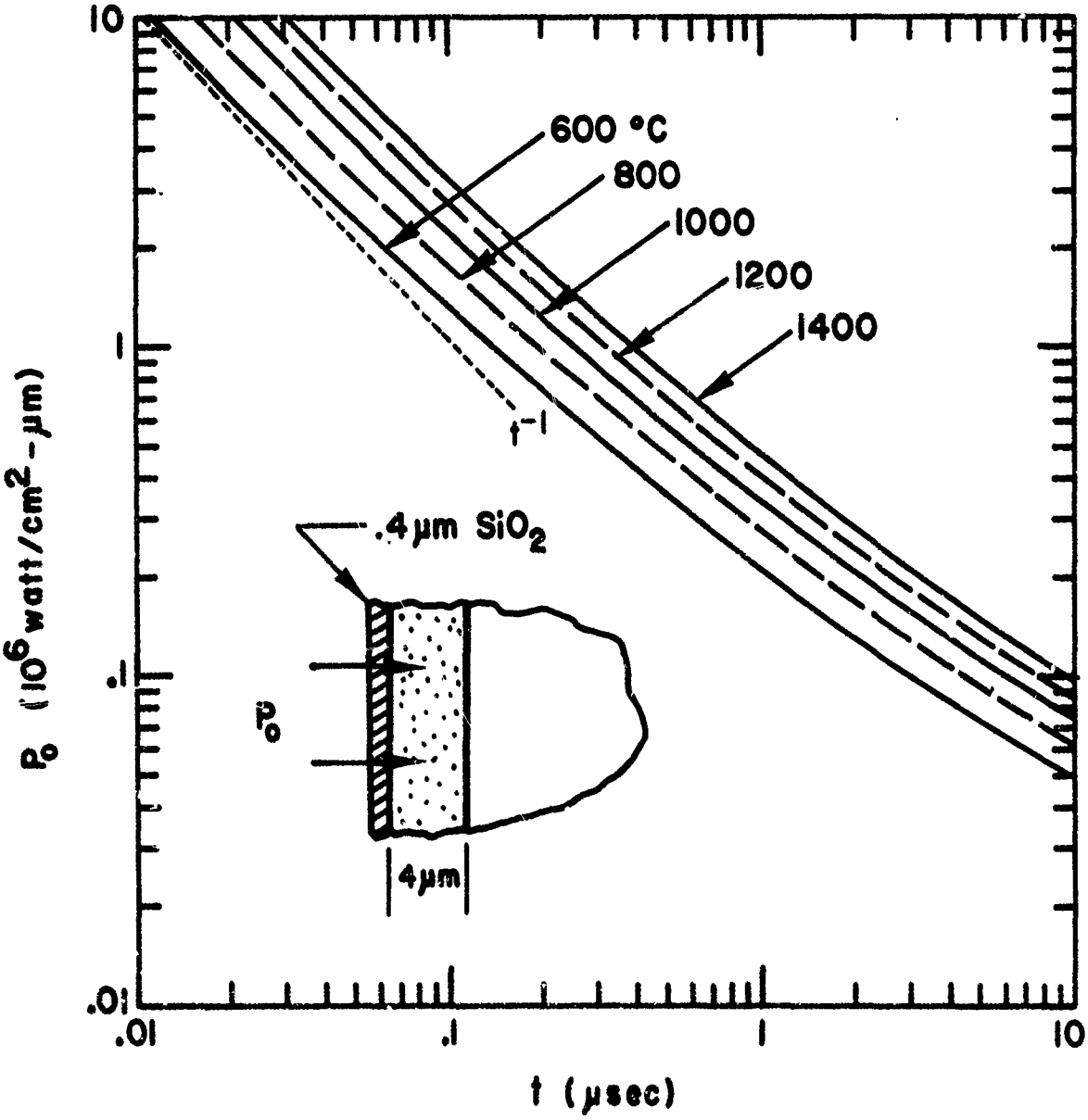


Fig. 19. Same as Figure 17, with the power applied to a 4 μm layer.

thicknesses of typical microcircuits.

Comparison of the three figures is obscured by the fact that the same power density is a different total power input in the three cases. The thicker the "active" layer the greater the total power input and the greater the temperature rise. For times less than 50 nsec the heating is nearly adiabatic and T_m does not depend on layer thickness. The three figures coincide in this range and the power density vs time curve varies as t^{-1} as shown in Figure 19. For longer times the curves would be expected to vary as $t^{-\frac{1}{2}}$ when heat transfer to the nearby silicon becomes appreciable. In Figure 17 the $t^{-\frac{1}{2}}$ assymptote has been reached, while for thicker layers the quantity of heat in the active layer if still large compared with the heat energy conducted away and the situation is still adiabatic - like in the time scale of the figures.

These last three curves are useful for estimating the temperature which a junction reaches when it goes into second breakdown. Unfortunately, the one-dimensional solution is a limitation. In an integrated circuit, the lateral heat flow into surrounding silicon is important when t is greater than about 100 nsec.

To summarize the presentation in this section, we can say that conduction is the only important mechanism of heat flow in integrated circuits. Although analytic solutions to the heat flows equation are available for a wide variety of relevant conditions, they cannot be applied with confidence when the temperature rise is large because of the variation in the thermal conductivity of silicon. Accurate answers are obtained by numerical methods using finite difference techniques. Solutions are best expressed in the form of power versus time relationships which can then be compared to

experimental findings.

In the next section a brief survey of second breakdown in semiconductor devices will be given. This will serve as an introduction to the experimental results presented and discussed in later sections.

SECTION III

A SURVEY OF SECOND BREAKDOWN IN INTEGRATED CIRCUITS

The effect in bipolar transistors commonly referred to as "second breakdown" was first reported in 1958 by Thornton and Simmons.[15] They suggested that it could be the cause of certain failures occuring in electronic circuits. It appears as a sudden transition to a low voltage, high current mode of operation accompanied by current constriction, which may or may not cause permanent changes in transistor characteristics or even catastrophic failure.

Although the term second breakdown was initially coined to refer to bipolar transistors operating in the normal, common-emitter configuration with the base terminal open, forward biased, or reverse biased, similar phenomena occur in diodes, both forward and reverse biased, MOS devices, microwave diodes, 4-layer devices, and, indeed, in all types of semiconductor devices. The very large number of papers published on second breakdown and related areas, the varied description of events which accompany second breakdown, and the proliferation of models and explanations for the phenomenon attest to its elusiveness and complexity. This section will describe briefly the historical development and features of the popular models, and some of the more recent experimental work.

As mentioned above, the first observations of second breakdown were in 1958. The previous year, Tauc and Abraham reported the thermal breakdown behavior of silicon p-n junctions.[16] These and later papers offered a host of explanations. The similarity between breakdown in junctions, transistors, and other devices led to proposals of a common underlying mechanism. These included the formation of a central microscopic melt,

a critical temperature, a critical current, a lateral thermal instability, a pinch effect, an instability due to space charge effects, and others. Special issues of the IEEE Transactions on Electron Devices, August/September and November, 1966, were devoted to second breakdown. In 1967 a comprehensive survey summarized and critically reviewed the theories in vogue at that time.[17]

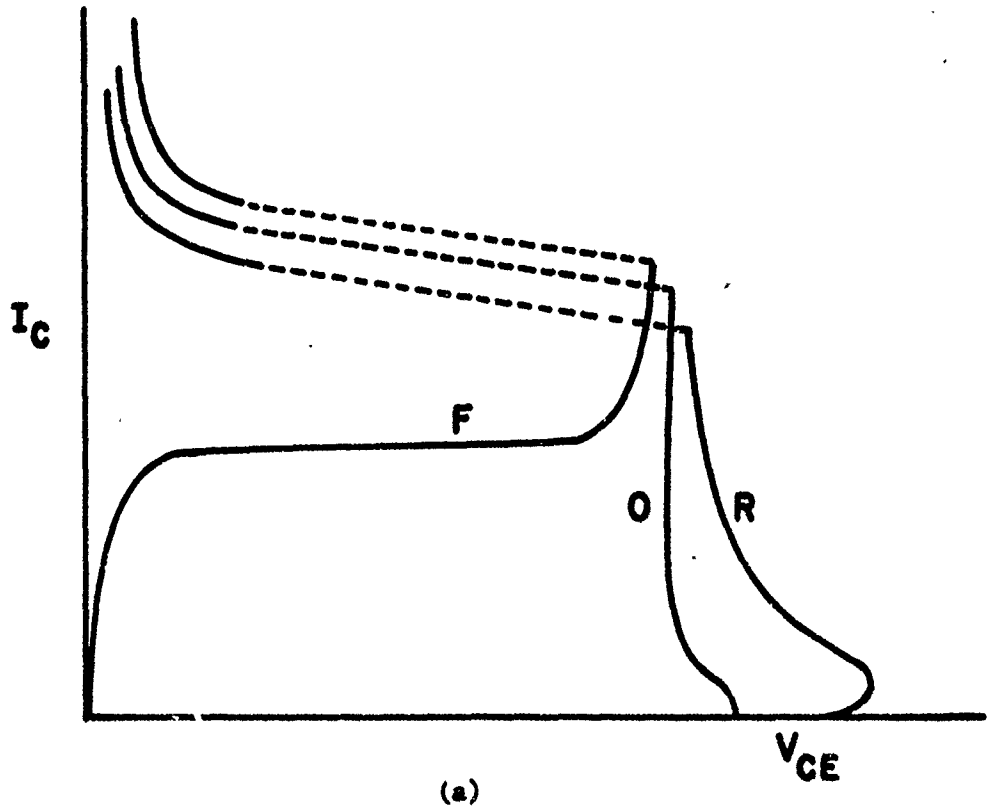
There are two common methods of observing the transition into second breakdown.* One is by direct observation of the volt-ampere characteristic on a curve tracer. Figure 20 shows typical curves for a transistor and a diode. The transistor is operated in the grounded emitter connection with the base either forward (F), open (O), or reverse (R) biased. Schafft[17] points out that a criterion sometimes used for second breakdown is that the voltage after breakdown is lowest for R and highest for F. The diode in part (b) of the figure is assumed to be operating in the reverse biased mode, although similar results are obtained for a forward biased diode.

In the second method, a voltage or current pulse is applied to the device under test. If the amplitude is great enough, after a certain delay time, which can be as short as nanoseconds, a transition to a new voltage and current level will occur. Prior to the transition there is often evidence of heating such as a rise in voltage or decline in current. After the transition the voltage and current levels are usually quite stable. Typical waveforms are depicted in Figure 21.

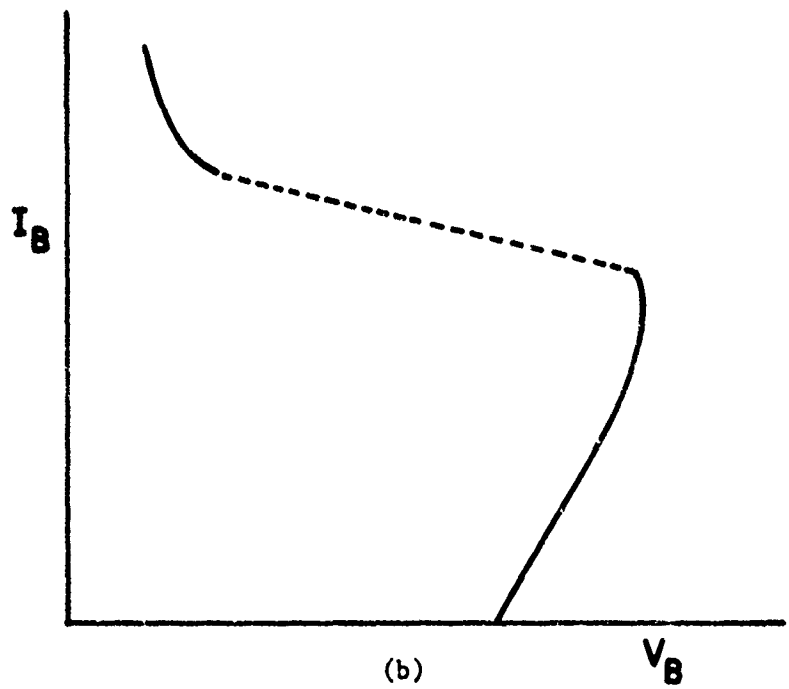
Following an excursion into second breakdown permanent changes in the device may be observed, depending on the structure, heat sinking, length

---

* During the remainder of this report the term second breakdown will be used to mean the sudden change to a low voltage, high current mode of operation of any semiconductor device.



(a)



(b)

Fig. 20. Second breakdown characteristics of semiconductor devices. (a) Transistors, showing breakdown under forward F, open O, and reverse R base biasing. (b) Diode breakdown in the reverse biased direction.

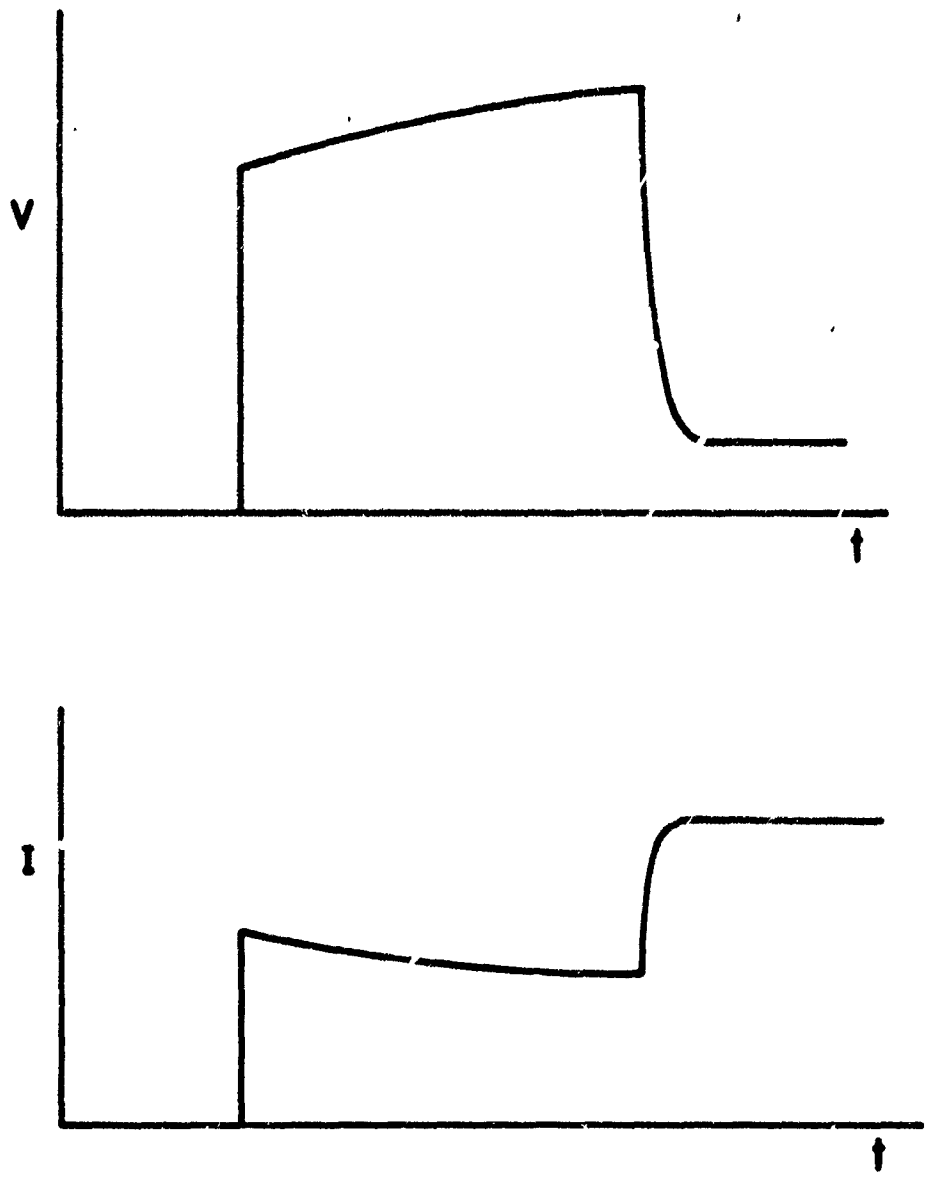


Fig. 21. Pulse voltage and current waveforms of a device stressed into second breakdown.

of time in second breakdown, and the power dissipation before and after the transition. A microscopic examination may reveal no observable damage, darkening of the contacts, a stress mark at a junction, "white streak" damage, or massive junction and metallization damage. Electrical changes may not be detectable, or may include increased leakage current, soft junction breakdown, changes in breakdown voltage, changes in beta, a decrease in delay time during subsequent pulsing, or the presence of a low resistance short circuit across the contacts.

For devices where the delay time depends on the amplitude of the pulse, it is clear that a thermal mechanism must be involved. The concept of an intrinsic temperature was proposed by Ford[18] and by Melchior and Strutt.[19] In this theory, once the junction temperature reaches a point where the lightly doped side of the junction becomes intrinsic, then further increases in temperature will cause a decrease in resistivity. The negative resistance results in an instability which causes a current constriction to form. The low resistance of the constriction causes the collapse of the junction field and the drop in terminal voltage.

Another version invokes a critical temperature at the point where thermal generation of carriers is able to satisfy the current requirements without the need for avalanche multiplication. When the device becomes hot enough the current concentrates at the point of highest temperature and the extinguishing of the avalanche produces the observed drop in voltage. If this is the case, the presence of defects should have an effect on the delay time. Investigations into this point have produced mixed results according to Schafft.[17] His conclusion is that second breakdown does not depend on imperfections for its existence but is enhanced by them.

If the idea of a critical temperature is valid, changes in ambient temperature should cause a difference in the delay time. Again the results are mixed as reported by Schafft.

In the microscopic central melt theory[20] it was postulated that a small portion of the junction melts during the transition to second breakdown. If the melt or "mesoplasma" were small enough it might not be detected during an analysis of the device. A number of other workers have investigated the temperature and voltage distribution in melted columns, the stability of melted regions, and the voltage-current relationships.

More recently a number of other important contributions to the work on second breakdown have been made. Unfortunately, the picture has not been simplified; instead, new evidence has been presented for the existence of several distinct mechanisms.

In the case of diodes, Bowers[21] and Muller and Guckel[22] have shown that space charge effects due to device current can cause negative resistance under certain conditions. Apart from the instability which accompanies negative resistance, the space charge will cause a change in voltage across the space charge region which then appears as a change in terminal voltage. With high current pulses the voltage is considerably higher than the junction breakdown voltage and it is important to know how much of the excess appears across the junction and how much is due to parasitic IR drops in the bulk material, leads, etc.

Hower and Reddi[23] have reported that second breakdown can be induced in n^+-n-n^+ diodes by a process of avalanche injection. The diodes were sufficiently small and the pulse power and pulse time such that the peak temperatures did not exceed 38°C above ambient. Under these conditions, if

second breakdown is observed, it is easy to argue that the effects cannot be thermal! They observed a negative resistance region at low currents which was attributed to space charge effects. At higher currents a discontinuity to a low voltage high current state was observed. This was correlated to the onset of avalanche multiplication at an n^+-n junction. Transistors fabricated on the same wafer with diodes were also found to have second breakdown characteristics caused by avalanche injection.

Second breakdown triggered by a critical current level was found by Tasca in both diodes and transistor junctions.[24] He did not attempt to fit his findings to any particular mechanism. On the other hand, he also reported that some of his devices failed by a thermal mechanism.

Chiang and Lauritzen[25] worked with small, heavily-doped, guard ring diodes. They monitored the light emission due to avalanche breakdown and incandescence at current constrictions. Thermal breakdown occurred at about 1100°C, which was also the temperature at which the lightly doped side became intrinsic. The injected carrier concentration at thermal breakdown was about two orders of magnitude less than the background doping concentration.

Fleming[6] and Wang[26] have argued the case for the existence of a critical temperature where the leakage current has increased enough to provide the total device current. Avalanche multiplication is no longer required and the avalanche is extinguished. This point of view is also advanced by Sunshine who has performed a series of detailed and comprehensive experiments on breakdown in silicon-on-sapphire diodes.[27]

Sunshine's work is an extension of work by Dumin[28] who had concluded that second breakdown was reached when the diode current density was

equal to the background doping concentration. Sunshine, using SOS diodes supplied by Dumin, devised careful experimental techniques which enabled him to ascertain the sequence of events which accompanied second breakdown. In the first place, he reexamined Dumin's work on light emission during avalanche breakdown, current constriction at the junction, and filament formation in the high resistivity region. He also assembled light transmission apparatus for determining the temperature of the silicon material based on the variation of the optical absorption coefficient with temperature. Finally, he used a laser in a stroboscopic technique which enabled him to observe temperature variations in space and time with repetitive pulses.

The diodes were made by growing a thin silicon film, either n or p type, about 1 μm thick on a sapphire substrate. p^+ and n^+ regions were diffused to form p^+nn^+ or p^+pn^+ diodes. Metal lands on the p^+ and n^+ regions served as contacts for probes. The diodes were about 50 μm wide, and the high-resistivity region was varied in length from 2 to 300 μm. Breakdown voltage depended on doping level, being typically 18 volts.

The dc experiments clearly showed the relationship between the diode volt-ampere characteristic and the current distribution in the diode. At low current levels avalanche light emission showed nonuniform breakdown across the junction. The silicon had a high dislocation density, but was otherwise free from defects. In spite of the non-uniform breakdown the volt-ampere curves showed a sharp, clean breakdown unlike that when microplasmas are present. As the current level increased the avalanche emission became more uniform and more intense, showing the levelling effect of the negative temperature coefficient of avalanche breakdown. At still higher currents the current began to constrict, usually near the center of the

diode where the temperature was greatest. At a level corresponding to the threshold current for the onset of negative resistance (second breakdown) the avalanche emission was extinguished in a portion of the junction. At the center of this region a small incandescent spot appeared. As the current was increased still further, the spot grew in size and intensity and avalanche emission ceased everywhere.

A constant current source was also used for pulse testing. When the pulse current was less than the threshold value the diode voltage would show the effects of diode heating. When the current exceeded the threshold value, two characteristic time intervals were noted: T_D, the delay time, and T_F, the filament formation time. When the current pulse was applied the voltage would initially increase due to heating. At time T_D the voltage reached a maximum value, then began to decline rapidly. After an additional time interval T_F the voltage reached a steady, sustaining value.

With the stroboscopic technique Sunshine was able to correlate these characteristic times with variations in current and temperature in the diode. At time T_D two things happened. First, avalanche emission ceased in a region near the center of the diode. Second, a hot spot formed at the center of the region, while portions of the junction adjacent to the hot spot began to cool. After careful investigation, Sunshine concluded that the extinction of the avalanche and the formation of the hot spot occurred when the local temperature was high enough to generate sufficient thermal current to supply the diode current density. Avalanche multiplication being no longer required, the voltage across the junction at that point would drop The drop in voltage would cause the nearby current to funnel into the region, resulting in a current constriction.

The constriction would elongate into a filament extending across the high-resistivity region during the time interval T_F if the growth conditions were favorable. This required that the power density be high enough to raise the temperature of the high-resistivity material to its intrinsic temperature.

Constriction occurred at junction temperatures of several hundred degrees Celcius. As the filament formed and became incandescent the peak temperature increased, reaching 700-1100°C before burnout occurred.

In the case of forward biased diodes the maximum temperature rise occurred near the center of the high resistivity region rather than at the junction. Filament formation again took place when the temperature reached the intrinsic value.

Sunshine's work has been extended and elaborated on by Budenstein, Pontius, and Smith.[29] Their conclusions generally agree with those of Sunshine. They used SOS diodes of various geometries and resistivities fabricated by the Autonetics Division of North American Rockwell.

The diodes were pulsed with a constant current source. At low pulse amplitudes the avalanche emission was low in intensity and beaded in appearance. There was little indication from the voltage waveform or from the transmitted light photograph of heating. At higher pulse amplitudes the avalanche was brighter and more uniform with evidence of heating at the junction and to a lesser extent throughout the high resistivity region. At still higher levels, at junction temperatures between 600 and 700°C, current constriction occurred at certain sites along the junction. The constriction sites appeared in the transmission photographs as dark spots along the junction, with avalanche emission ceasing in a region 1 to 5 mils on either

side. The dark spots usually occurred at the same places at each pulse. The number and spacing was determined by the pulse amplitude and duration. The short, high amplitude pulses produced the largest spot density.

The appearance of a current constriction was accompanied by a slight, often imperceptible drop in voltage. As the filament spread across the high resistivity region, the voltage decreased slowly. When the filament finally bridged the entire region catastrophic failure occurred as a molten column about 1 μm wide formed within the current filament. At this point the voltage dropped abruptly to about one-half its peak value. The damage produced during this final stage was visible and also appeared as a deterioration of the volt-ampere characteristic. Budenstein suggests that what Sunshine refers to as second breakdown is the initial drop in voltage during the nucleation and filament formation and that the final, abrupt change in voltage accompanied by device damage did not occur. Thus there is a somewhat different interpretation of second breakdown between the two groups of researchers.

Many other features of Sunshine's work were corroborated by Budenstein et al. Nucleation of a filament was found to occur when the thermally generated carriers were able to sustain the current. The filament spread into the high-resistivity region as the temperature was raised to the intrinsic temperature.

In forward biased junctions or semiconducting material without junctions a similar sequence of events took place. Heating occurred in the high resistivity bulk material or around the probe contacts. A current filament formed, and when it finally bridged the contacts a molten filament formed and the voltage dropped abruptly. In some reverse biased diodes it

was possible to initiate a filament in the high resistivity region before constriction at a junction site took place.

The summary by Budenstein et al., provides us with a fairly clear picture of second breakdown in junctions. Two stages of filament growth are associated with second breakdown. In the first stage current constriction occurs and a filament grows, accompanied by a relatively slow, gradual drop in voltage. In the final stage the filament has bridged the contacts and a melt column forms, accompanied by a precipitous drop in voltage. These two stages cannot always be clearly distinguished in the voltage waveform.

In the case when a reverse-biased junction is present nucleation of a filament occurs at the junction when the thermal generation of carriers is sufficient to supply the current requirements. Otherwise, a filament forms when the temperature of the bulk material reaches the peak of the temperature-resistivity curve.

Junction defects play only a minor role in the nucleation of filaments. The position of junction dark spots depends mainly on the temperature distribution. When a melt channel has been formed, a subsequent breakdown does not usually superimpose a second melt channel on the first.

Although only thermal mechanisms were considered it seems reasonable that other carrier injection mechanisms would bring about second breakdown whenever the injected current density was equal to the local junction current density.

This summarizes the present understanding of second breakdown in semiconductor junctions. A great deal of progress has been made in recent months. The following section will present the results of pulse testing of integrated circuits, showing the relation with the theories of this section.

SECTION IV

EXPERIMENTAL RESULTS

In the experimental part of this program, several different integrated circuits were stressed into second breakdown with single pulses. The object was to determine the relationship between threshold power and pulse duration. The pulse voltage and current waveforms and the device damage will be related to the second breakdown models discussed in the preceding section. Recall that second breakdown is defined here as the transition to a low-voltage high-current state. In the case of reverse biased junctions this transition almost always resulted in permanent electrical changes and visible damage. The rare exceptions will be pointed out. In the case of transistors, the results are somewhat more complex.

A. Testing Procedure

The integrated circuits were pulsed into second breakdown by the application of approximately constant current pulses ranging from 100 nsec to 10 msec long. The device under test was connected to a Cober Model 604A pulse generator. Amplitude of the pulser is continuously variable in three ranges from 0 to 1500 volts open circuit. Currents up to 6A are available, with a peak power of 9 KW. Rise time is adjustable from 20 nsec to 100 nsec. Pulse width is determined by an external pulse generator and can be varied from 50 nsec to dc. Output impedance for the 0-150 volt range, where most testing was done, is about 40 ohms. The generator was found to be reliable, simple, and convenient to use, and was able to supply sufficient power for nearly all test conditions. Voltage across the DUT was displayed on one channel of a Tektronix 555 dual beam scope while the other beam monitored

the current by means of a Tektronix clamp-on current probe. All testing was of a single pulse nature. A Datapulse 101 pulse generator operating in the single pulse mode triggered the Cober pulse generator and fixed the pulse width while simultaneously triggering the oscilloscope sweep. A Polaroid camera recorded the current and voltage waveforms of each pulse.

Figure 22 shows the experimental arrangement. Most tests were performed at room temperature. In some cases the ambient temperature was changed by immersing the device in liquid nitrogen or mounting it on a hot stage. In some instances the DUT was mounted under a microscope with a video camera attached. The results of single or repetitive pulsing could be viewed on a monitor and simultaneously recorded on an Ampex VR7800 video tape recorder.

Junctions were stressed in both forward and reverse biased modes. Transistors were tested collector-emitter, collector-base, and emitter-base. Electrical characteristics were checked on a curve tracer before and after testing. The devices were delidded for visual inspection after testing. In a few cases the devices were decapped and photographed prior to testing.

The usual procedure was to select a pulse width and gradually increase the pulse amplitude until second breakdown occurred. Second breakdown was detected either from the voltage-current waveform or from changes in the electrical characteristics displayed on a curve tracer. To rule out the possibility that the initial, sub-threshold pulses could somehow alter the device's susceptibility to second breakdown, several junctions were pulsed 50-100 times or were pulsed repetitively at low prf for several minutes. The voltage-current waveforms were found to be identical before and after, the electrical characteristics showed no evidence of change, and

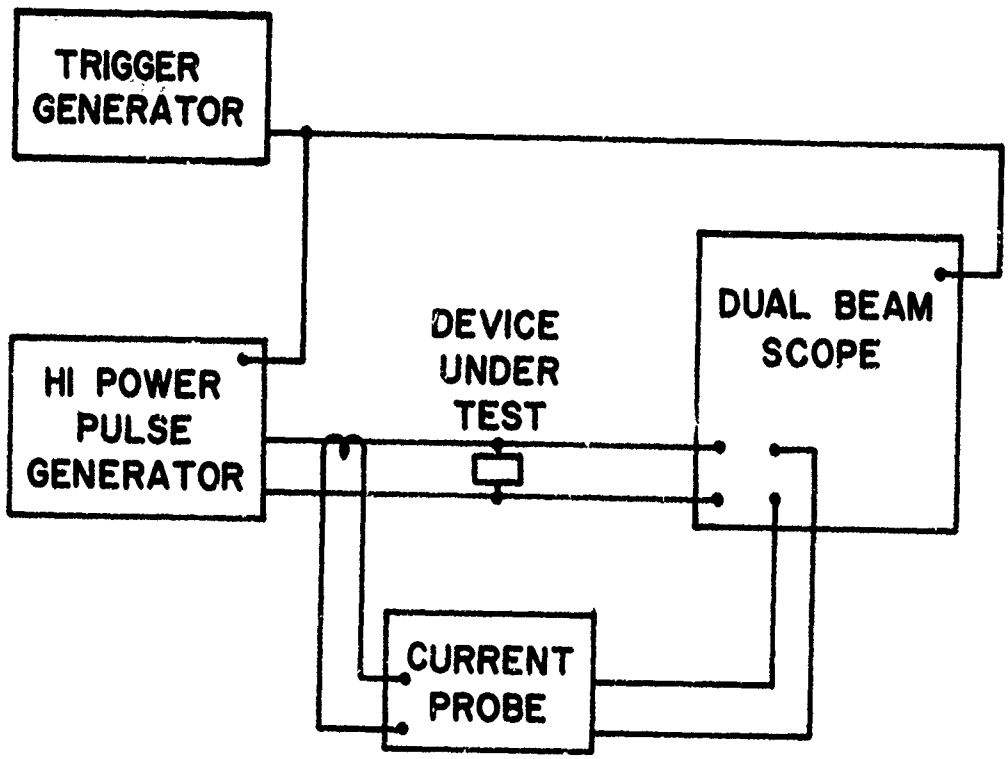


Fig. 22. Experimental arrangement for pulsing devices. The current and voltage waveforms were recorded by a Polaroid camera.

the second breakdown characteristics were subsequently found to be virtually identical with those of other junctions on the same chip.

After stress-testing, most devices were examined visually and photographed. Some circuits were examined in greater detail by probing, angle-lapping, cross-sectioning, and etching.

B. Devices Tested

All devices were in monolithic integrated circuit form, either commercially available or fabricated by standard techniques by major semiconductor manufacturers. Tests were performed only on parts where single junctions or transistors could be isolated from the rest of the circuit. This placed some restriction on the selection of devices; nevertheless, a reasonable selection of manufacturers and fabrication technologies was made.

A list of devices is given in Table I. With the exception of the core drivers, all circuits were low power devices. Up to 20 parts of each kind were tested, each part containing up to 6 available junctions and transistors.

C. Results

1. Part A

This device was tested more exhaustively than any of the others. It is a dual, 3-input expander, mounted in a 10 lead TO-5 can. The circuit diagram is given in Figure 23. The connected emitters and collectors preclude the possibility of testing individual transistors unless the metallization is scratched open. If the collector-base junction of one transistor is pulsed in the reverse biased direction, instead of observing BV_{CBO} on that transistor we observe BV_{CEO} on one of the adjacent transistors plus the

TABLE I

TEST DEVICES

Part	Manufacturer	Type	Devices Tested	Comments
A	1	Dual 3-input gate expander	Emitter-base diodes	Diffused collector technology
B	2	Special circuit	Diodes, transistors	Special test circuit, DTL technology
C	3	DTL expandable dual 3-2 input NAND gate	Diodes	Glassivated
D	4	DTL dual expandable NAND gate	Diodes	Dielectrically isolated
E	3	Dual 4 input diode expander	Diodes	Dielectrically isolated
F	4	Quad core driver	Transistors	Dielectrically isolated

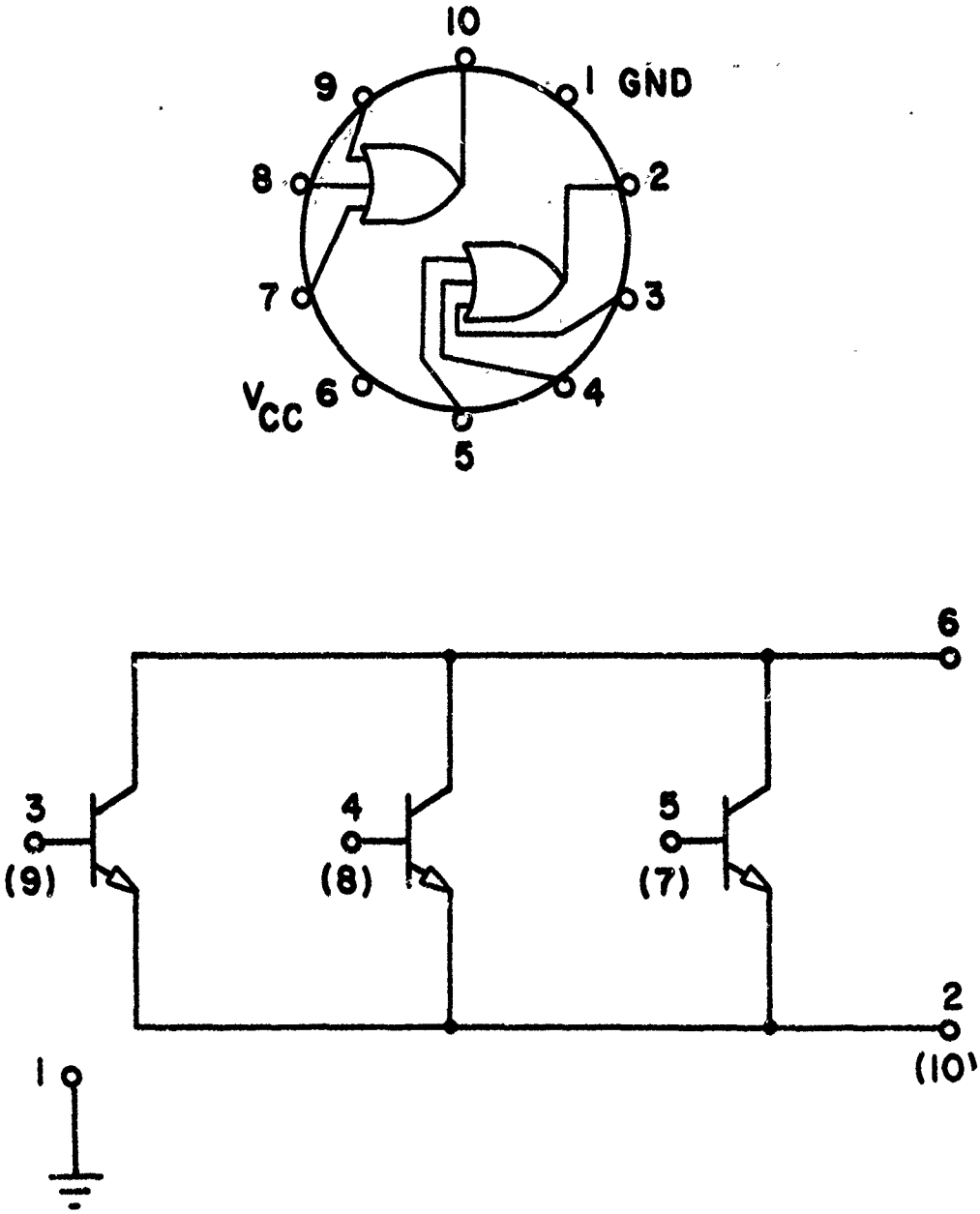


Fig. 23. Pin connections and schematic diagram for Part A.

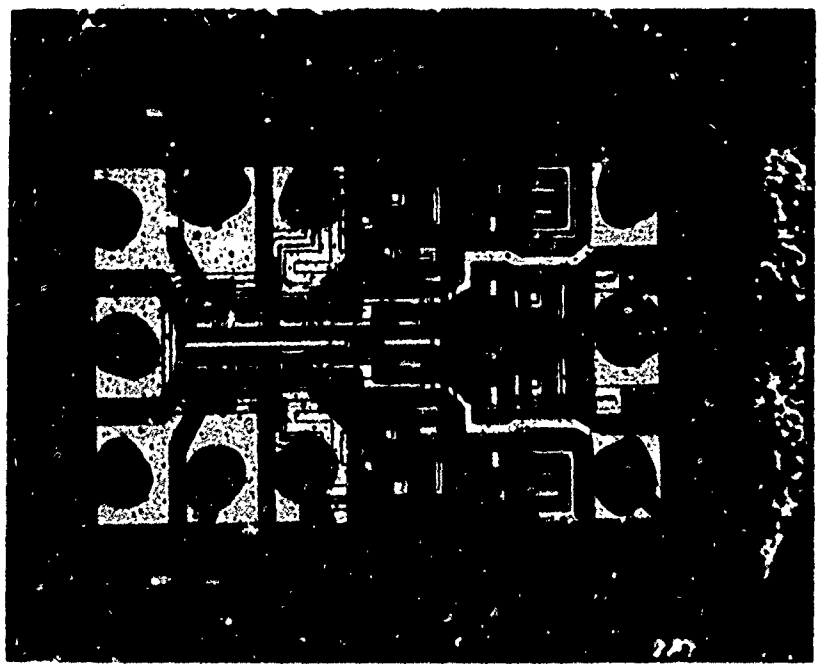
emitter-base breakdown voltage of the transistor being tested. Consequently only the emitter-base junctions were tested.

Photomicrographs of the circuit appear in Figure 24. The chip size is 34 x 54 mils x 7 mils thick. Details of the emitter-base region are shown in part(c) of the figure, where each small division is 2.5 μm. Note the misregistration. The distance between the emitter junction and the base contact was found to vary from 5 to 17 μm in the 17 parts examined. A correlation was found between this spacing and the threshold power for second breakdown. Also note that some of the contact windows are only partly covered with aluminum, the emitter contacts on one side and the base contacts on the other. This has implications regarding the effect of contact resistance and contact alloying on second breakdown threshold energy. No relationship was noted on these devices.

Angle lapping and cross sectioning several samples yielded the dimensions shown in Figure 25. There is no epitaxial layer, the collector regions being formed from a deep diffusion. The oxide thicknesses were of the order of 0.15 μm over the emitter regions, .5 μm over the base diffusions, and 0.63 μm over the remaining portions of the die. Resistors were probed on 3 chips to determine the base resistivity. It was nearly identical on all 3 chips, averaging 120 ohms per square.

The electrical characteristics of all devices were quite uniform. The breakdown voltage of the emitter-base diode was 6.4 volts with no evidence of soft breakdown, instability, etc. At 77°K the breakdown voltage decreased to 5.4 volts, while the forward voltage drop at 2 ma increased from 0.80 volts to 1.14 volts. BV_{CEO} was 7 volts, BV_{CBO} was 24 volts, and collector-substrate breakdown voltage was 42 volts. BV_{CBO} decreased by

(a)



(b)

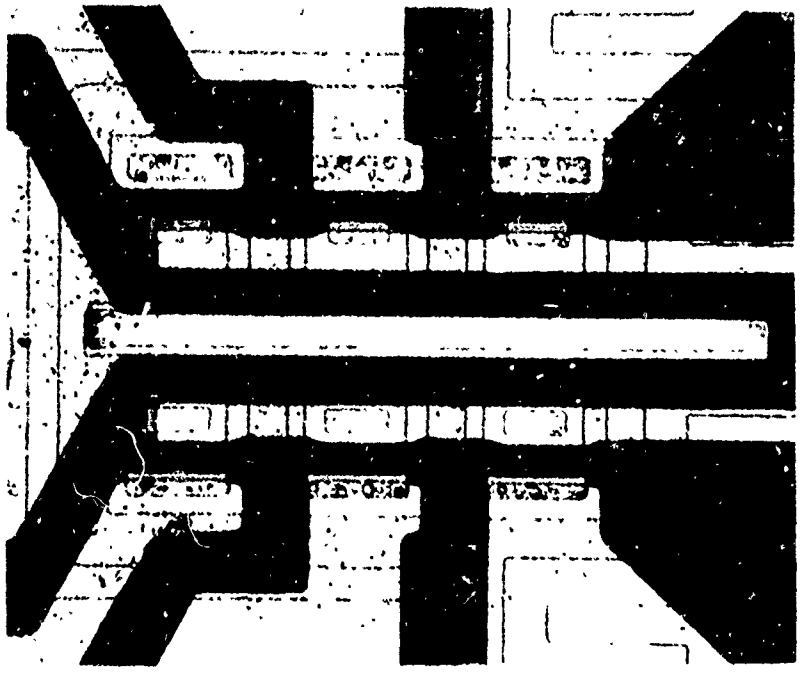


Fig. 24. Part A, dual 3-input gate expander. (a) Overall view of chip. (b) View of transistors tested.

(c)

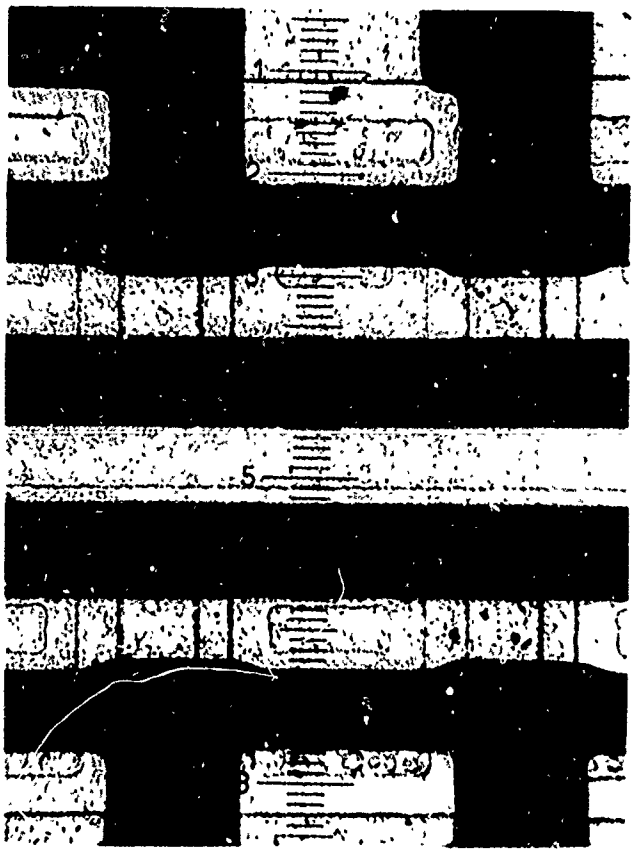


Fig. 24 (cont.) (c) View of emitter-base junctions. Note differences in junction-base contact distance, and metallization covering only part of some contact windows. Each small division is 2.5 μm.

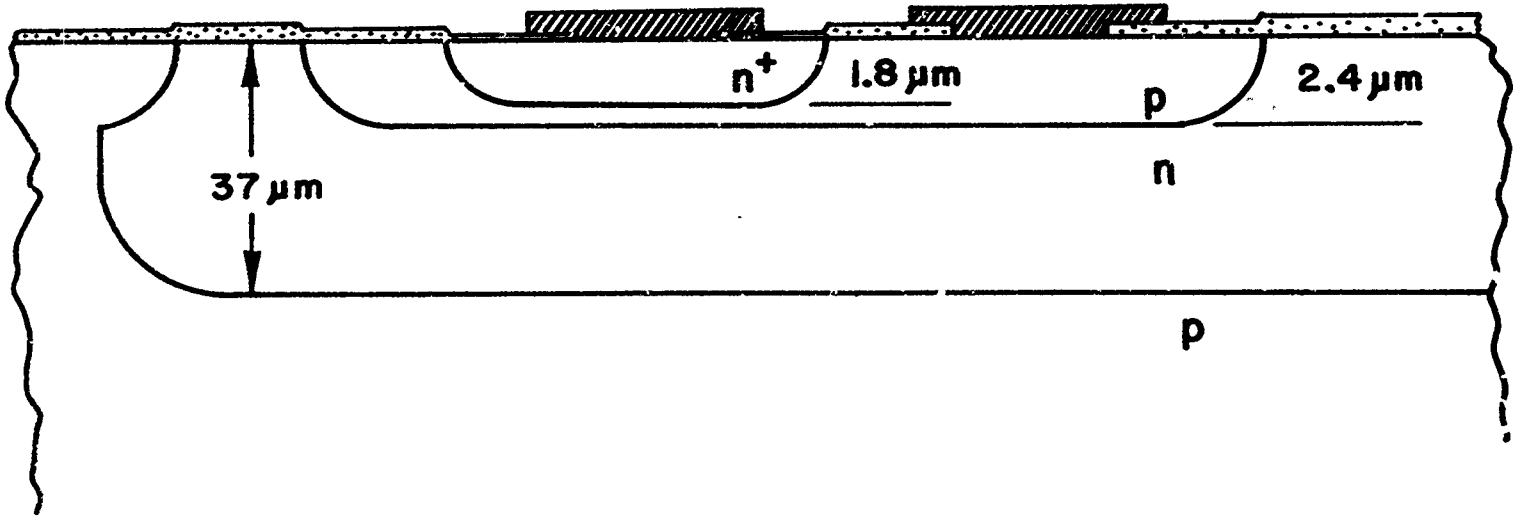


Fig. 25. Cross-section view of Part A. Chip fabrication was by the diffused collector process. Dimensions are not to scale.

5 volts at 77°K while the collector-substrate breakdown decreased by 6 volts.

As explained earlier, each junction was checked on a curve tracer first. Then the device was connected to the Cober pulse generator, a pulse width selected, and the device pulsed at increasing power levels until second breakdown occurred. Typical waveforms are shown in Figure 26 for reverse biased diodes at room temperature. Waveforms at -196°C and 200°C were similar.

The upper waveform in each case is the voltage across the junction and the lower waveform is the current. The second breakdown transition is readily apparent in each waveform. In parts (e) and (f) of the figure, the current is off scale after second breakdown. The pulse risetime is about 20 nsec. Current and voltage are roughly constant both before and after second breakdown. Heating effects are noticeable during the first part of the pulse as the voltage rises and the current falls. Eventually the slope becomes zero and then changes sign. This is an indication that second breakdown is imminent. In part (d) of the figure, the discontinuity in the slope is particularly evident in the voltage waveform about 1.5 μsec before second breakdown. This "glitch" was noticeable in most cases, appearing as a gradual change in slope in the shorter pulses (a), (b), (c), and as a "hesitancy" in the longer pulses (e) and (f). If the pulse was terminated after the glitch appeared but prior to second breakdown, no damage to the diode could be detected. Subsequent pulses at the same power level did not induce second breakdown. The voltage and current waveform remained the same, although the glitch sometimes shifted to a slightly earlier time in the pulse.

The interpretation of second breakdown by Rudenstein et al., seems to apply here. It is reasonable to ascribe the discontinuity in the voltage

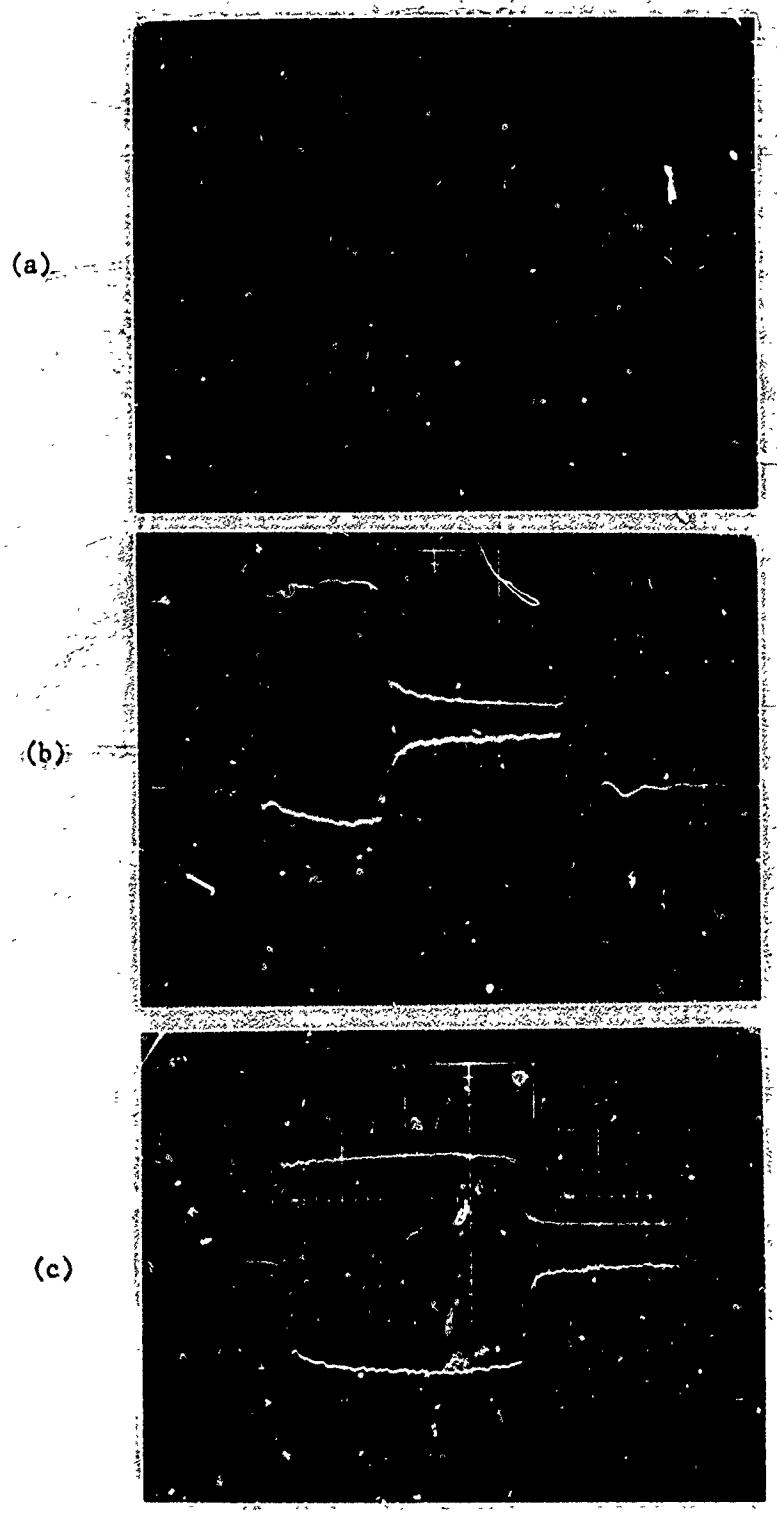


Fig. 26. Pulse waveforms for Part A in the reverse bias mode. (a) Upper waveform 10 v/cm, lower waveform 200 ma/cm, time axis 100 nsec/cm. (b) 5 v/cm, 200 ma/cm, 200 nsec/cm. (c) 10 v/cm, 200 ma/cm, 500 nsec/cm.

(d)

(e)

(f)

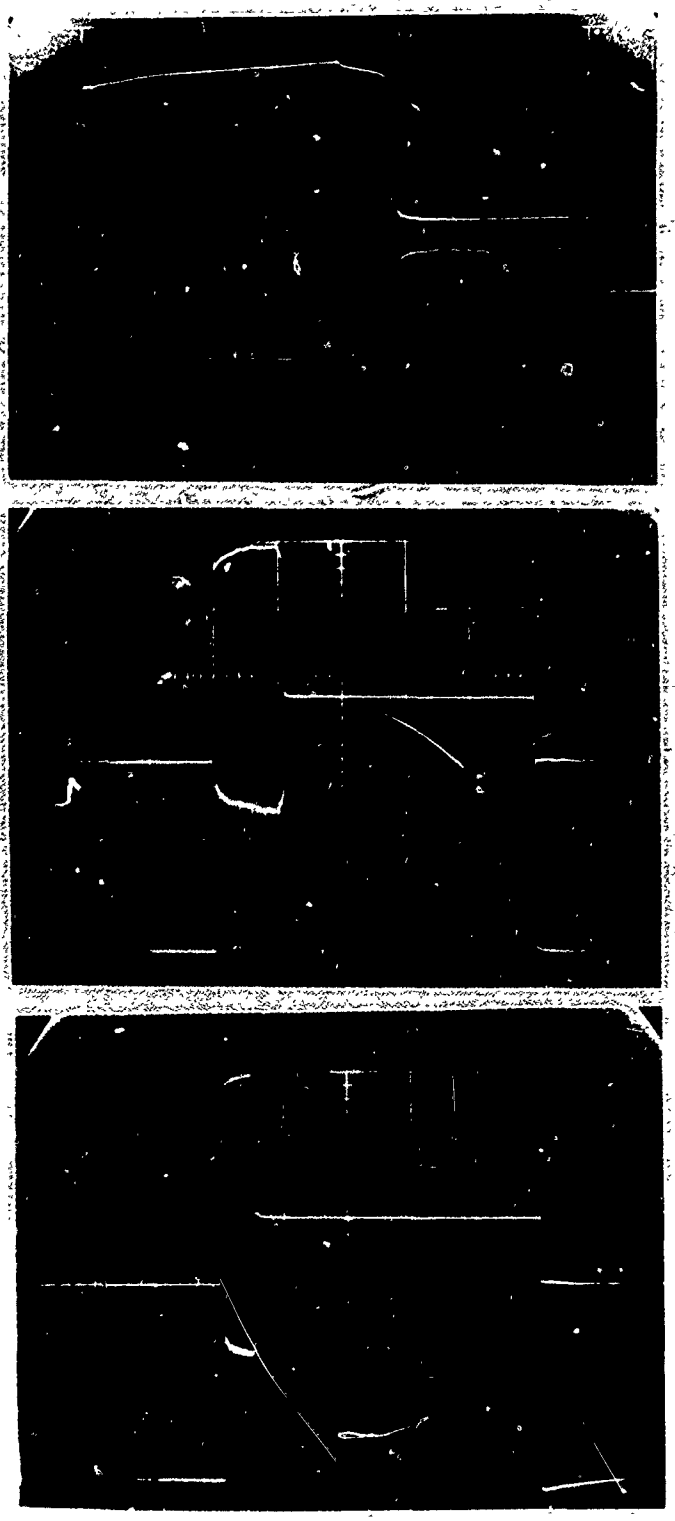


Fig. 26 (cont.). (d) 5 v/cm, 200 ma/cm, 2 μsec/cm. (e) 5 v/cm, 100 ma/cm, 20 μsec/cm. (f) 5 v/cm, 100 ma/cm, 200 μsec/cm.

to a change in the conduction process. But what kind of a change? The fact that there is no damage means that it is reversible. The transition to second breakdown always caused damage in these diodes, hence must be associated with the formation of a microscopic melt. What stage lies between normal avalanche breakdown and melt formation? The only conclusion is current constriction and filamentation, exactly as in the case of SOS diodes.

TABLE II

DIODE BREAKDOWN CALCULATIONS FROM FIGURE 26(c)

		Before Breakdown	After Breakdown	
Average	V	15	5	volts
Average	I	300	580	ma
Average	P	4.50	2.90	watts
Time	t	1.87	1.20	μsec
Energy	E	8.41×10^{-6}	3.48×10^{-6}	joules
Resistance	R	50	8.6	ohms

Table II lists the quantities of interest calculated from the waveforms of Figure 26(c). The average power is taken as the average current times the average voltage. The diode resistance is the ratio of average voltage and current. This is important when related to the generator output impedance (38 ohms in this voltage range). As the diode shifts into the second breakdown mode its resistance changes to a lower value. If this shift causes the diode to be more closely matched to the generator impedance the power supplied to the diode will increase. For test device A the diode was nearly matched prior to second breakdown. After second breakdown occurred there was a mismatch and the power decreased. This is an important considera-

tion when evaluating damage to the device.

Of primary interest is the relationship between the delay time before second breakdown and pulse power. This is plotted in Figure 27.* The numbers near the data points indicate the distance in microns between the emitter junction and the base contact window (or base metallization if the window is not entirely covered). The dashed lines connect points from junctions on the same side of a single chip. It is clear that junctions with the greatest spacing have a higher threshold power. Unfortunately, the distance parameter could not be controlled and there is not enough data for a quantitative relationship with the other variables.

Seven junctions were pulsed to failure in the forward direction. For this condition second breakdown was more difficult to detect from the voltage-current traces, particularly for short pulses. The junctions were checked on a curve tracer after each pulse to determine when permanent damage took place.

Waveforms of three junctions are shown in Figure 28. Part (a) shows two successive pulses of short duration superimposed. After the first pulse there was no observable damage so the amplitude was increased slightly. The voltage and current of the second pulse are slightly higher until nearly the end of the pulse when the voltage changes slope and becomes less than the voltage of the first pulse. After the second pulse both electrical and visual damage were observed.

Part (b) of the figure shows three pulses applied to the same junction. Heating becomes more and more pronounced as the amplitude is

* When pulse rise time and transition time to second breakdown are not negligible, for example in the case of very short pulses, the pulse width is taken between 75 per cent amplitude points on the voltage waveform prior to second breakdown.

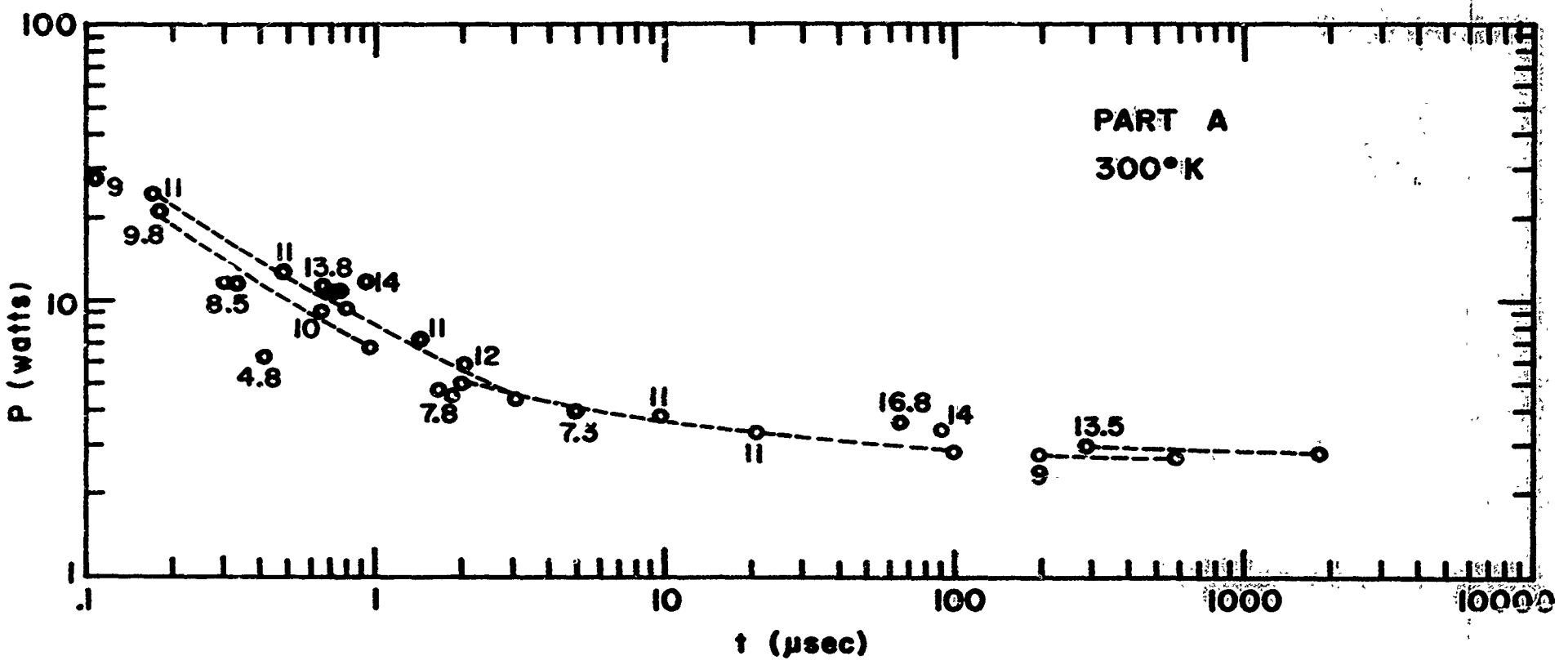


Fig. 27. Threshold power as a function of pulse width for reverse-biased emitter junctions of Part A. Numbers beside the data points are the junction to base contact widths in μm. Dotted lines connect points from junctions on the same side of a chip.

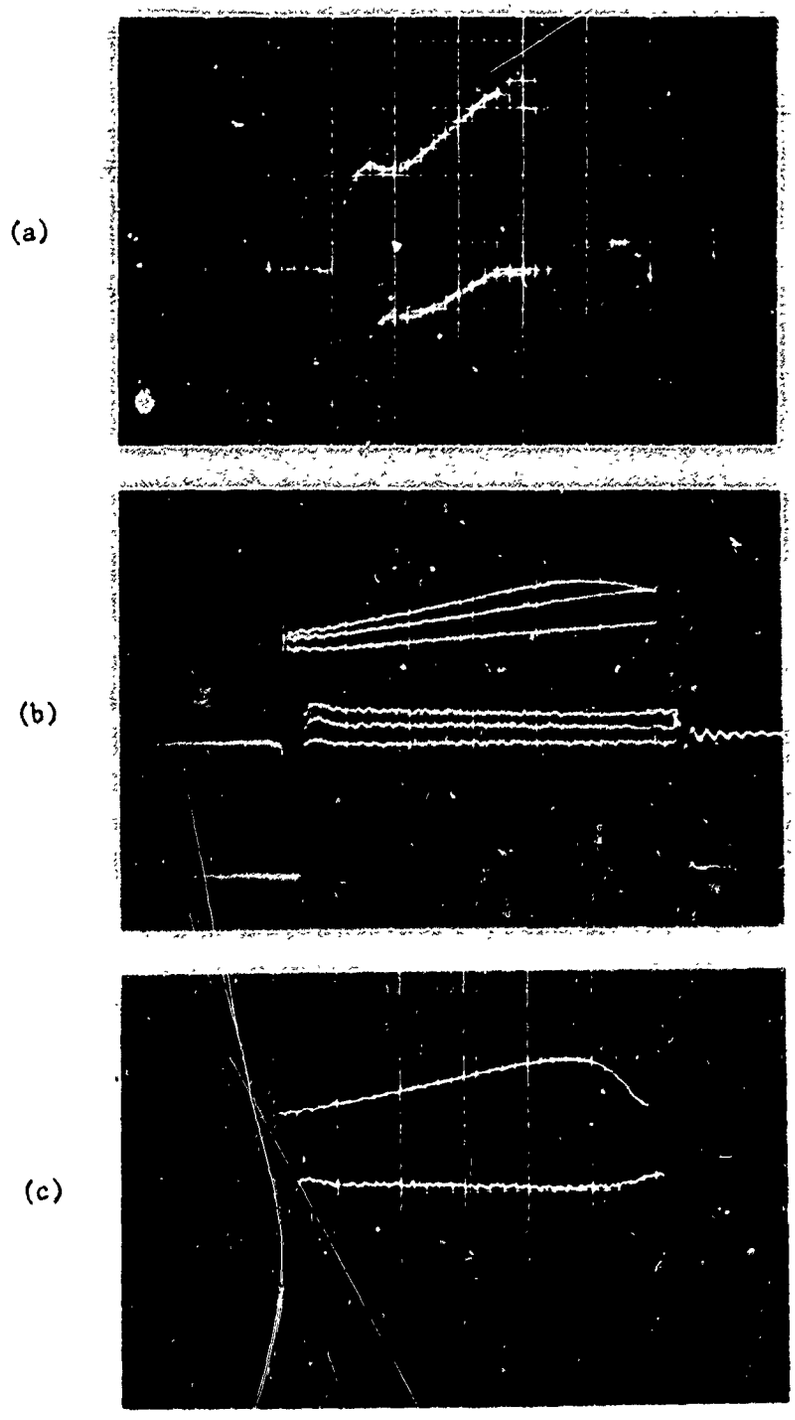


Fig. 28. Forward bias waveforms for Part A. (a) 10 v/cm, 2 A/cm, 100 nsec/cm. (b) 5 v/cm, 500 ma/cm, 500 nsec/cm. (c) 5 v/cm, 500 ma/cm, 500 nsec/cm.

(d)

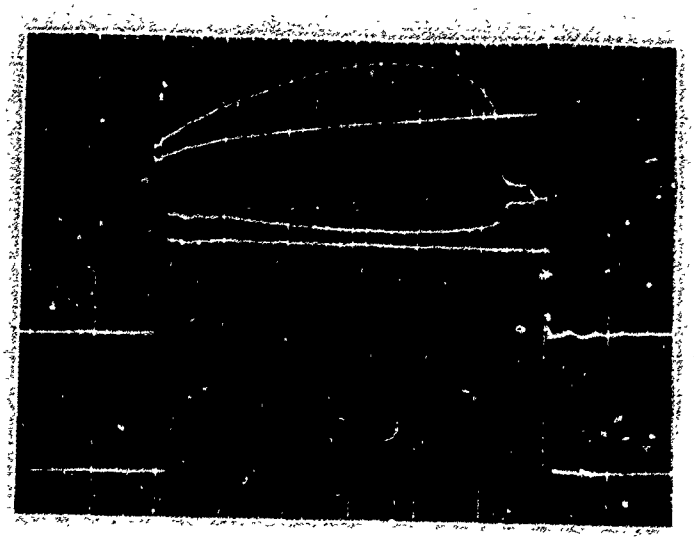


Fig. 28 (cont.). (d) 2 v/cm, 200 ma/cm, 10 μsec/cm.

increased. During the third pulse the voltage reaches a maximum value then decreases. No damage could be observed. Finally the pulse in part (c) of the figure was applied. Near the end of the pulse the voltage shows a greater drop and the rise in current is more apparent. This pulse caused permanent damage.

Part (d) shows two pulses of much longer duration. During the first pulse the voltage rises monotonically while the current remains constant. During the second pulse the voltage rises rapidly, reaches a broad maximum, and then drops abruptly in two steps. Permanent damage was found.

As in the case of reverse biased junctions the change in slope of the voltage is an indication that second breakdown is imminent. However, damage can occur without the precipitous drop in voltage which reverse biased junctions show. Budenstein et al., noted similar behavior in their diodes. Their findings indicate that the current distribution and heating are uniform up to and including the change in slope. Not until second breakdown is imminent does a filament form. Since the advent of permanent damage is not well defined in the short pulses of Figure 28, one concludes that the peak voltage point must occur at a very high temperature; when filamentation finally does take place the formation of a melted column does not require much additional temperature rise and the voltage does not change much. For long pulses the temperature and current distributions are less uniform and the transition to a melted column would be more noticeable.

Junctions were also "zapped" at liquid nitrogen temperature and at 200°C, about 10 junctions at each temperature. Usually three junctions on the same side of a chip were tested at the three ambient temperatures and with the same pulse width to keep as many variables as possible the same.

Voltage-current waveforms were much the same at each temperature. Threshold levels were shifted, particularly at 77°K and at longer pulse lengths. At pulse widths less than 1 μsec the change in threshold power was small, especially when the ambient temperature was changed to 200°C. This is further evidence that filamentation and second breakdown occurs at very high temperatures.

The results of varying the ambient temperature are plotted in Figure 29. The curves are drawn for an average junction-base contact width of 10 microns. Data points are omitted to avoid confusion except for the forward biased cases where all seven junctions had nearly identical geometries.

The three reverse bias curves have not yet reached a t^{-1} asymptote at the shortest pulse width. This is because the devices are small, and in the reverse bias mode the power dissipation is highly localized; the heat conducted away is an appreciable fraction of the total energy in the small heat-producing volume, so adiabatic conditions are not approximated with 100 nsec pulses. In the forward bias mode the heat is produced throughout a much larger volume. The amount of heat which is conducted away in 100 nsec is a smaller fraction of the total heat dissipated and the situation is more nearly adiabatic.

One circuit was used to investigate the effect of heat sinking. The device was decapped and half the junctions were zapped at different pulse widths. Then a drop of epoxy was placed on the chip. The difference in threshold power for pulses up to 5 μsec long was less than 10 per cent. This means that the silicon dioxide layer is a relatively good thermal barrier for these pulse times.

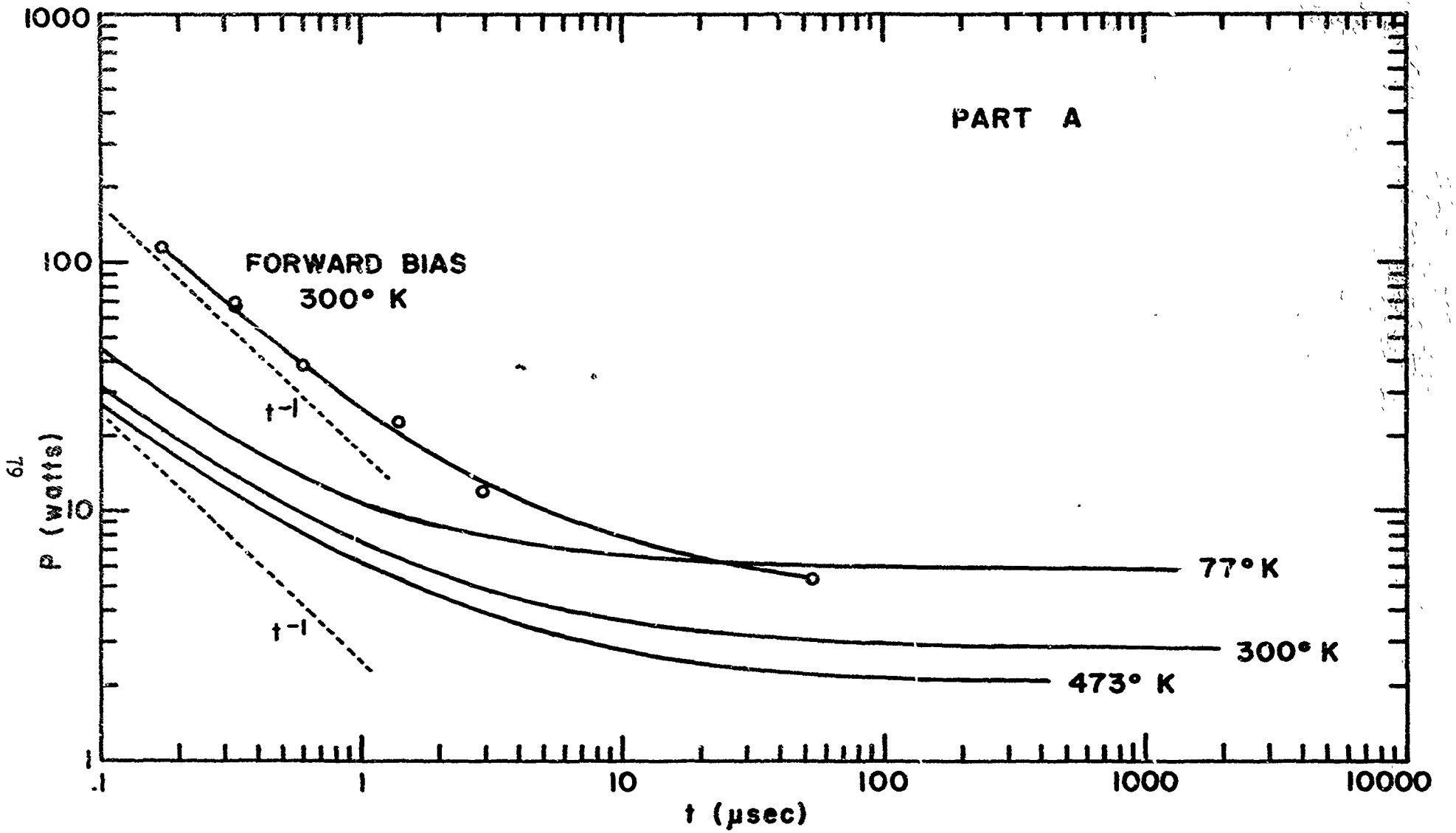


Fig. 29. Threshold power vs time curves for Part A. The three lower curves are for reverse-biased junctions with junction-base contact spacing of 10 μm.

Damage always occurred during second breakdown in reverse biased diodes, and in forward biased diodes whenever an easily discerned second breakdown transition occurred. The most sensitive indicator of damage was a change in the reverse breakdown characteristic of a diode. The forward characteristics were unaffected unless massive damage had occurred, in which case the diode appeared to be shunted by a resistor. The damage typically appeared as a change in the knee of the curve at breakdown, with the rest of the characteristic unaffected. The characteristics of several damaged junctions are shown in Figure 30.

It should be pointed out that a criterion of minimum detectable change is being used here. Damage does not necessarily imply a malfunction in circuit operation. The diodes in Figure 30 would perform satisfactorily in most circuit applications. The criterion of detectable damage however is more precise, independent of application, and is a limiting case.

Visual damage correlates roughly with the time spent in second breakdown and the energy dissipated during that portion of the pulse. Junction damage was visible in all but two cases. In these two cases the time spent in second breakdown was short - less than 20 nsec - but the stress marks were revealed when the metallization and oxide were removed and a dash etch applied for a few seconds. It should also be noted that visible damage might occur after the pulse has been terminated, for instance when heat from a high temperature region diffuses to a contact and causes delayed melting or contact damage.

Examples of damage in the reverse bias mode are shown in Figure 31. One of the junctions in part (a) shows aluminum in the melt track, with damage to the metallization on both emitter and base. Visible aluminum or

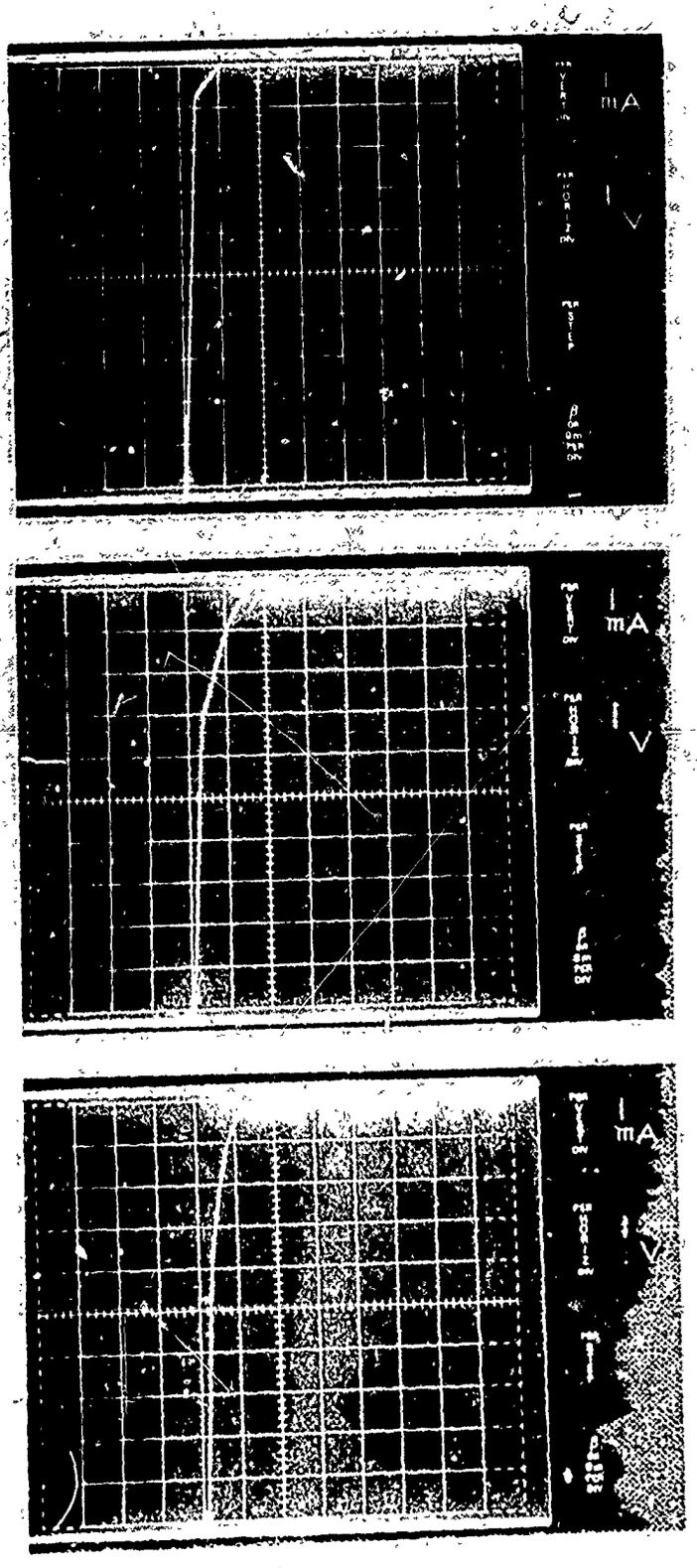
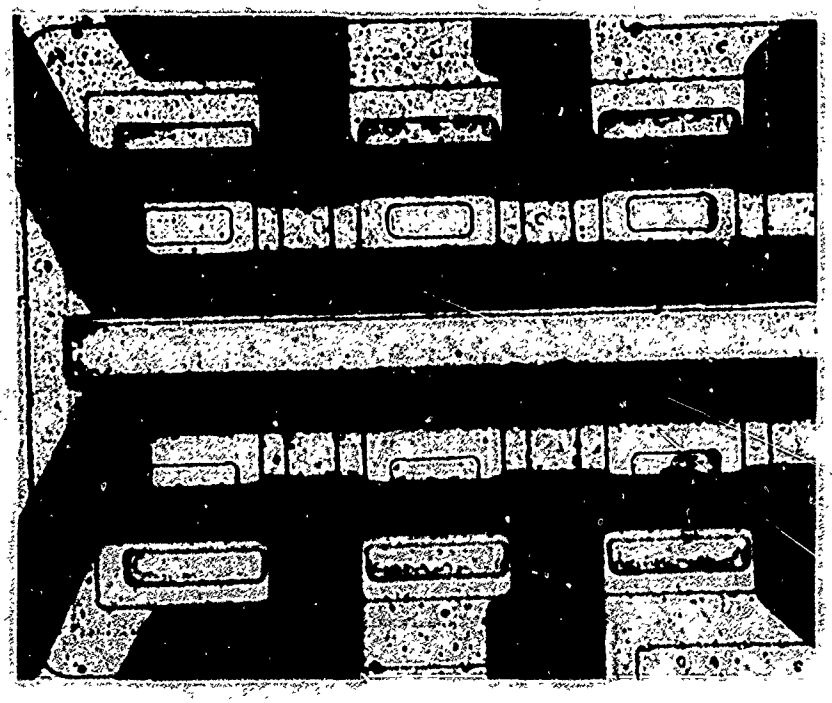


Fig. 30. Reverse bias breakdown characteristics showing damage due to second breakdown.

(a)



(b)

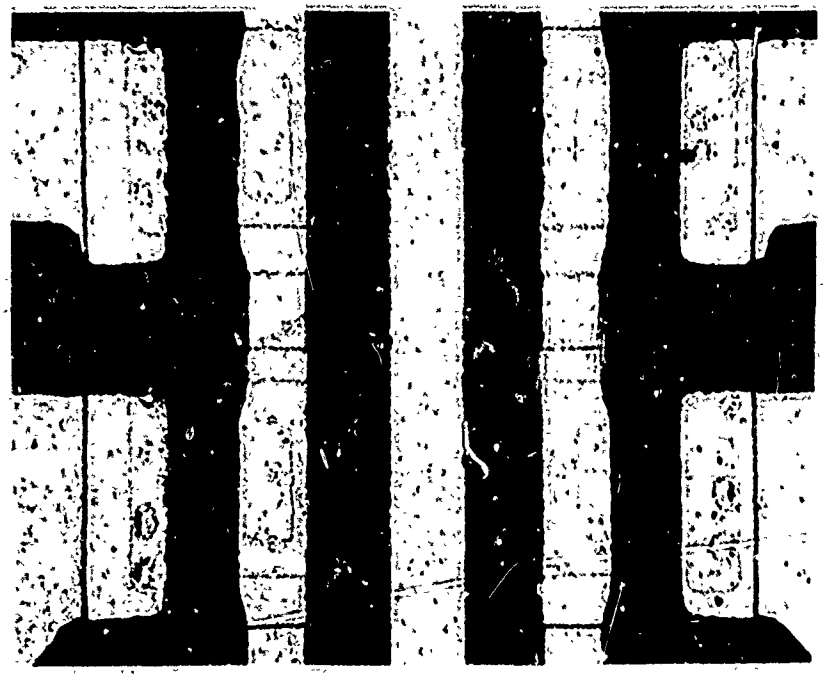
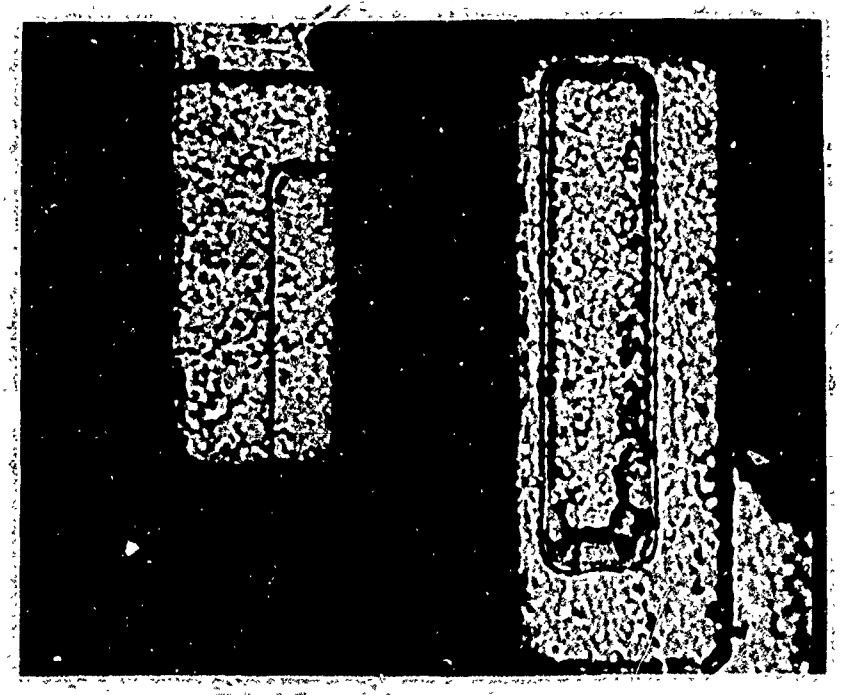


Fig. 31. Damage due to reverse bias pulses. In each case damage extends from the junction to the base contact.

(c)



(d)

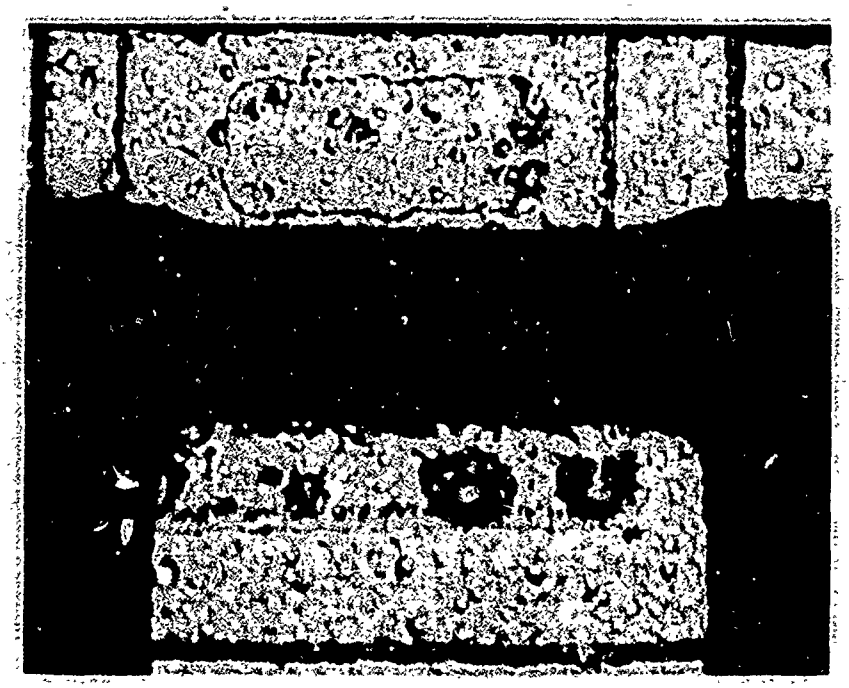


Fig. 31 (cont.).

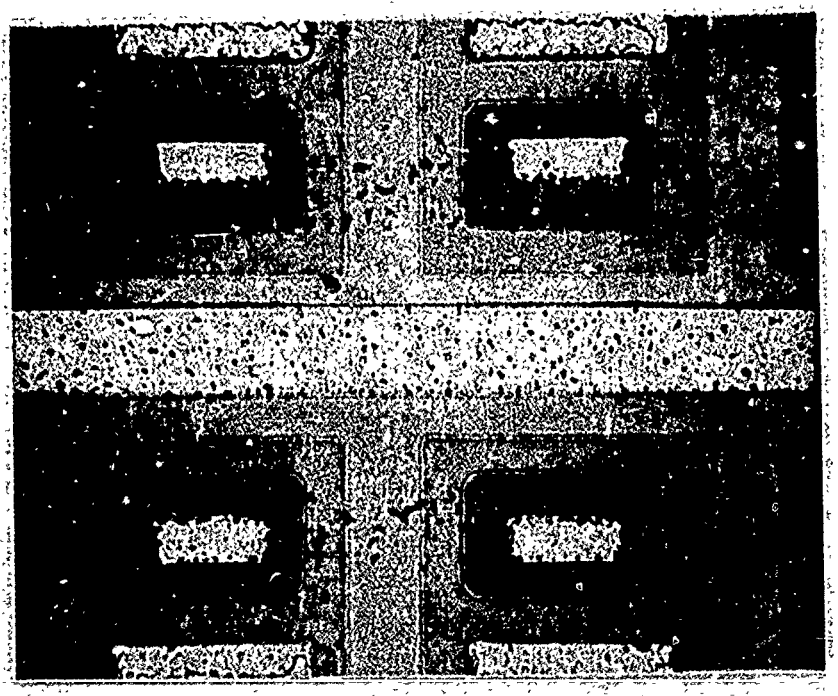
"white streak" damage was rarely seen on these devices. The other zap marks in part (a) are typical. Usually there is a pronounced mark at the junction with a tail extending toward the base contact. If the aluminum does not cover the base window entirely, or if the aluminum is removed, the track is seen to terminate at the base contact. There is usually no evidence of damage extending to the emitter contact.

In part (b) of Figure 31 the junction with the most damage was accidentally pulsed in the repetitive mode. In cases where this occurred the damaged region was often mushroom shaped with the cap extending laterally along the junction and well into the emitter contact. The stem extended to the base contact where metallization damage was apparent. The other stress marks in part (b) are typical. Parts (c) and (d) show two typical zaps at high magnification. In both cases the zaps terminate in the base window, but do not extend to the emitter contact.

In rare instances, damage was not apparent under the microscope. In Figure 32 one of these cases is shown. In part (a) the metallization has been removed. The device in the lower left has a very faint track in the base region. The device in the upper right has no visible damage. In part (b) of the figure the emitter and base oxide have been removed. This enhances somewhat the track at the lower left. In part (c) a Dash etch has been applied for a few seconds. In the upper right of the photomicrograph a faint mark is now seen at the corner of the junction. Thus it seems certain that second breakdown in these devices is always accompanied by electrical damage and detectable visual damage.

In some devices the zap marks extended well into the emitter contact area. This was the exception rather than the rule and usually happened

(a)



(b)

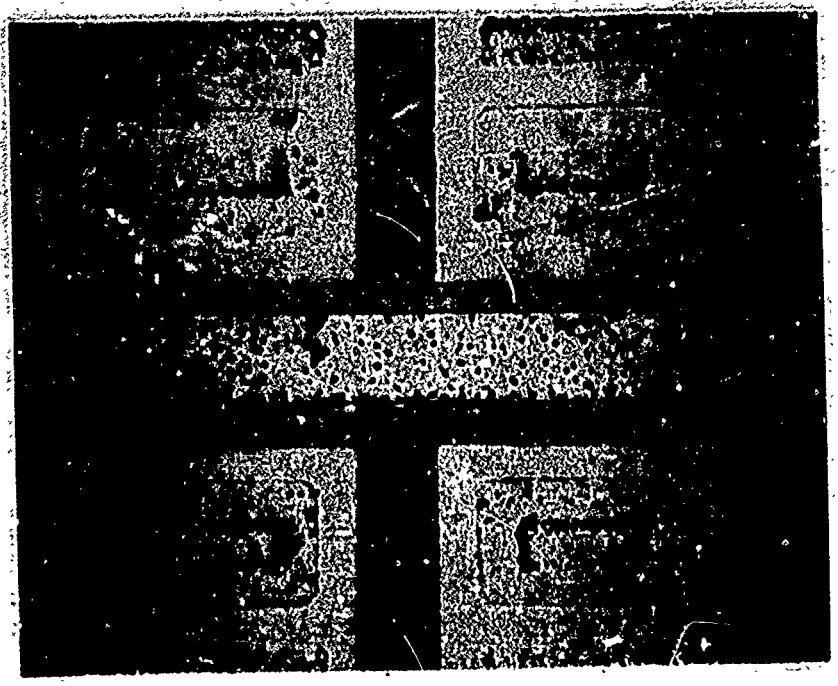


Fig. 32. Search for junction damage. (a) Metallization removal. No damage visible at junction at upper right. (b) Oxide etch. Still nc visible damage at upper right.

(c)

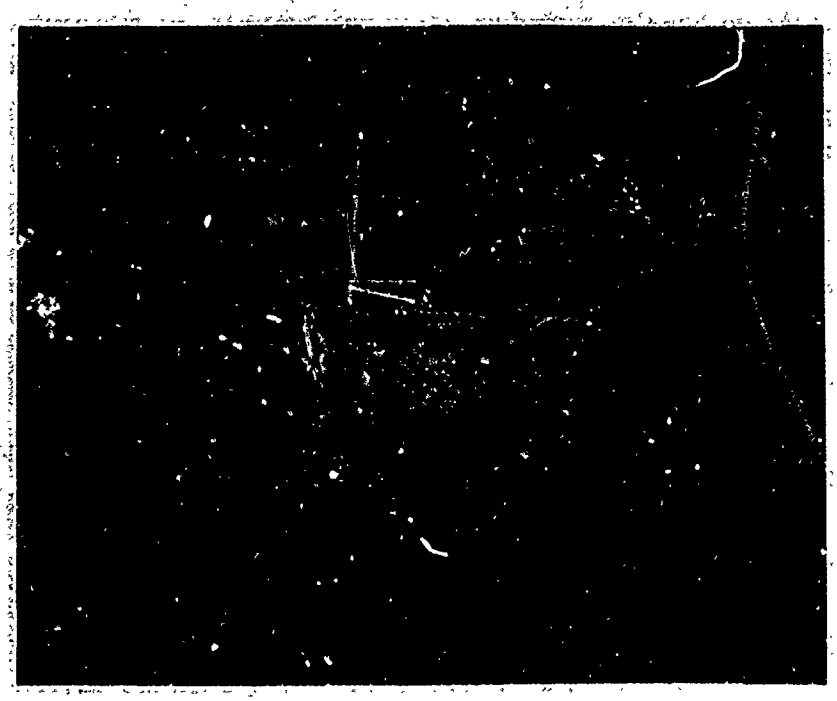


Fig. 32 (cont.). (c) Dash etch. A faint mark now appears at junction at upper right. A faint track extending to the base contact could be discerned under the microscope. The other stress marks clearly extend from junction to base contact.

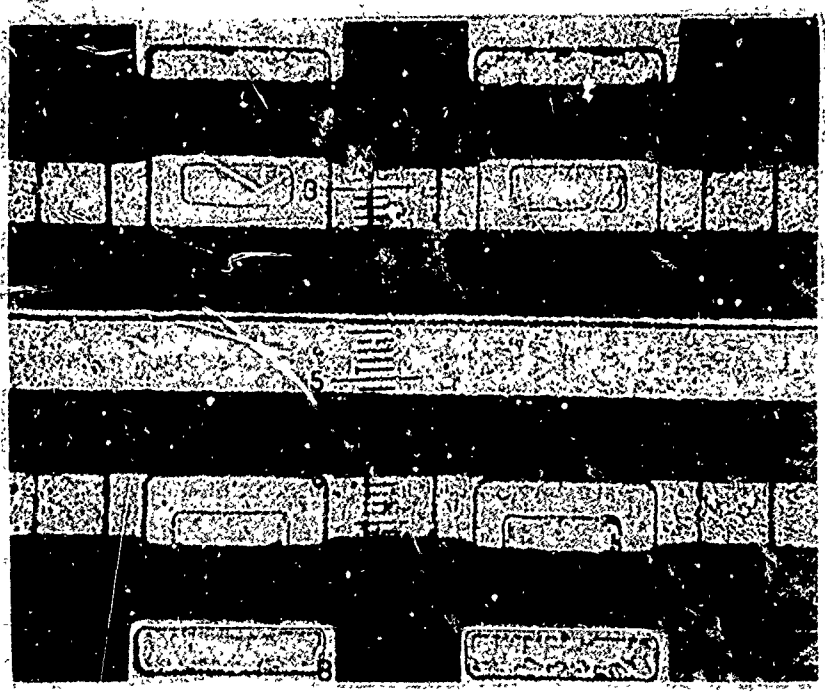
when the excursion into second breakdown had a long duration. Thus it appears that in reverse biased junctions the damage first takes place at the junction, spreads quickly into the base to the base contact, and finally extends through the emitter to the emitter contact.

Six forward biased junctions are shown in Fig. 33. In the first microphotograph, part (a) of the figure, several interesting features appear. The device at the upper left appears to be damaged only near the base contact. The device at the upper right has only barely discernable damage The damage to the diode at the lower left is normal, while the device on the lower right seems to be damaged only in the emitter region.

In the second microphotograph, part (b) of the figure, the metallization has been removed. This reveals some details not apparent in part (a). Finally in part (c) the oxide has been removed. All of the devices show at least faint tracks extending from the base contact to the junction.

In the device at the lower right corner the damage extends all the way to the emitter contact. The pulse which produced this damage is the one shown in Fig. 28(d). Referring to this previous figure, note the final short transition which occurs after second breakdown has taken place. In a later part of this report it will be shown that there is a strong correlation between the presence of the small voltage drop in the pulse waveform and damage extending from the junction to the emitter contact.

(a)



(b)

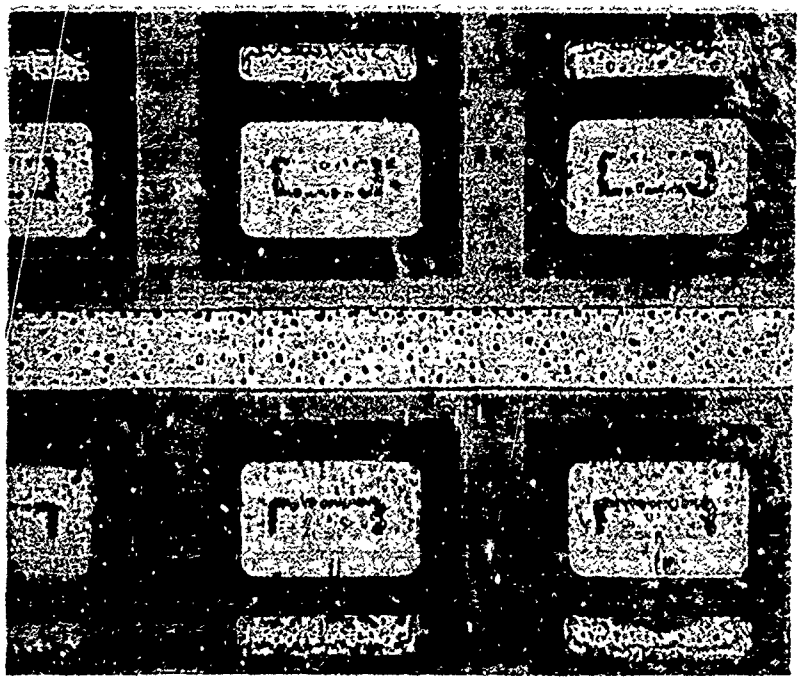


Fig. 33. Damage to forward biased junctions. (a) Photomicrograph of the junctions. (b) Metallization removal.

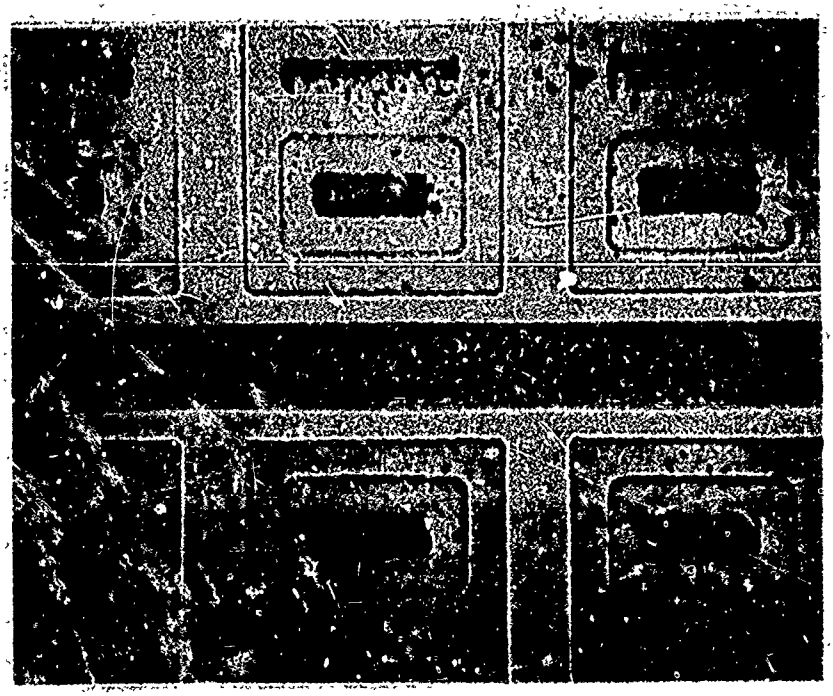


Fig. 33 (cont.). (c) Oxide removal. All junctions show damage extending from the base contact to the junction.

2. Part B

This circuit was designed for a previous test program. A number of parts were still available and have been utilized in the present study. Several components were fabricated on the chip. The circuit connections for the components tested in this program are given in Fig. 34. The diode was pulsed in both forward and reverse biased directions. The transistors were tested base-emitter, collector-base, and emitter-base. Unfortunately, it was not possible to remove the 1K resistor across the emitter-base terminals without opening the package and scratching open the metallization.

The chip is mounted in a 14-pin metal flat pack. Photomicrographs of the die are given in Fig. 35. Part (a) is an overall view. The chip measures 44 x 58 x 8 mils. Fabrication is by the planar epitaxial method with a buried subcollector, using standard 10 MHz DTL technology.

The diode is located at the edge of the chip, pins 8 and 10 the cathode and anode, respectively. There is a metallization stripe exactly over the active edge of the junction. Three transistors are accessible, with emitters connected in common to pin 1. The transistors are of roughly the same dimensions but with slight differences in the geometry. The differences can be seen more clearly in part (b) of the figure. The transistor near the middle of the die has a somewhat greater junction to base contact distance (recall that this made a significant difference in the threshold power of Device A, Fig. 27). The other two transistors have emitter windows of slightly different area and location within the emitter region. Details of the diode and one transistor are more easily seen in the last two photographs in the figure. The small divisions are 6.25 μm.

Other details are shown in Fig. 36. The epi layer is 10 μm thick.

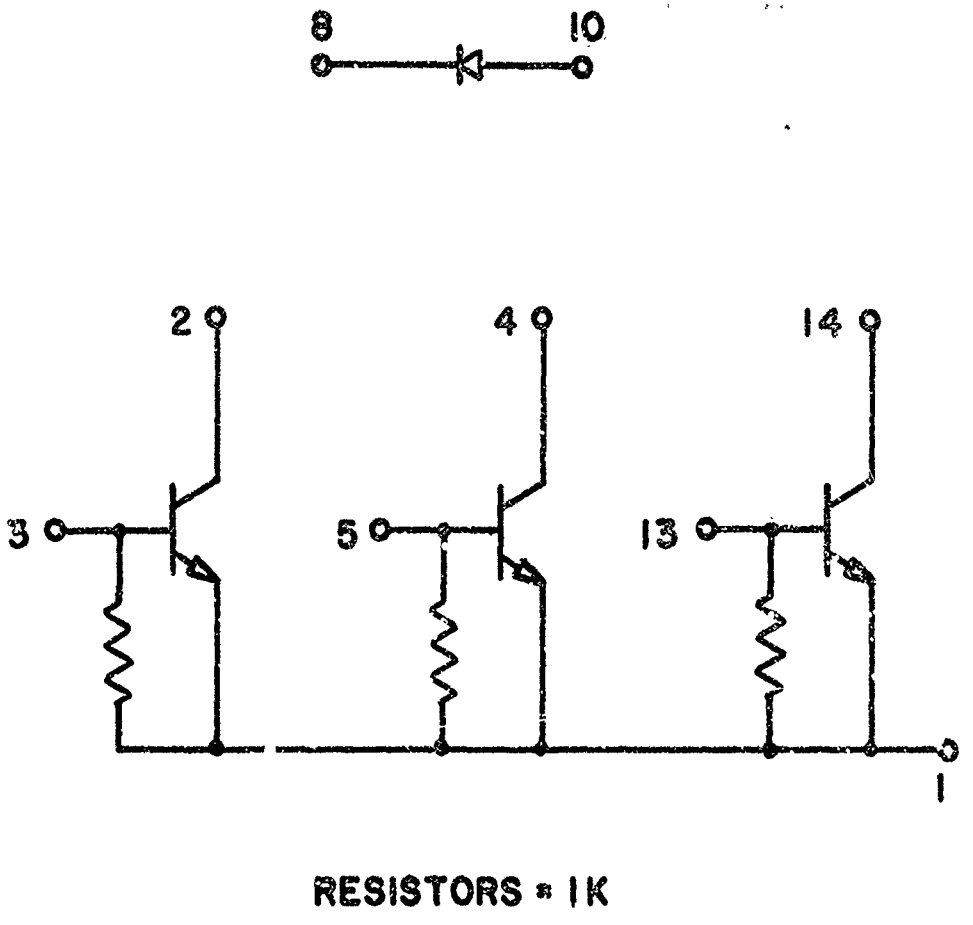
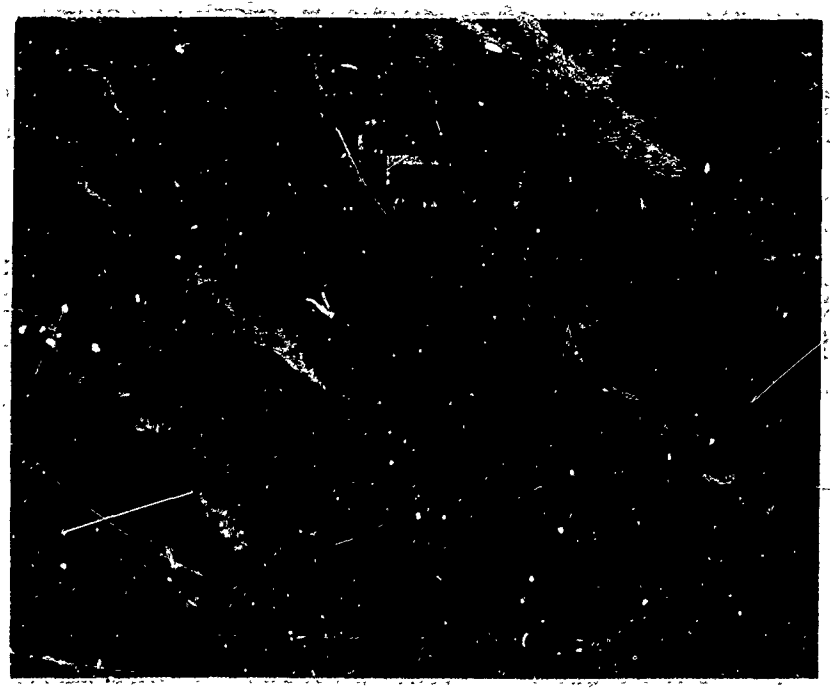


Fig. 34. Schematic diagram for Part B.

(a)



(b)

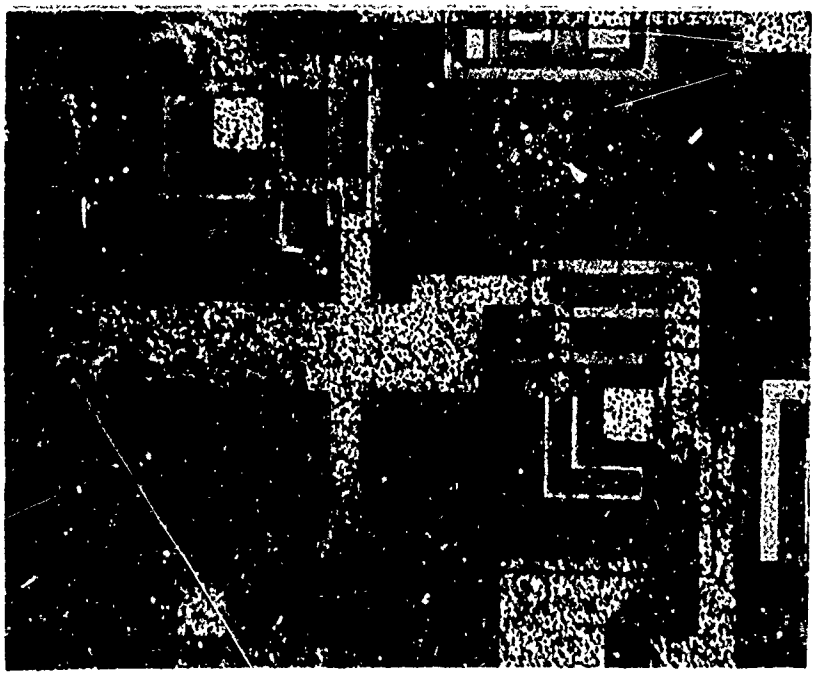
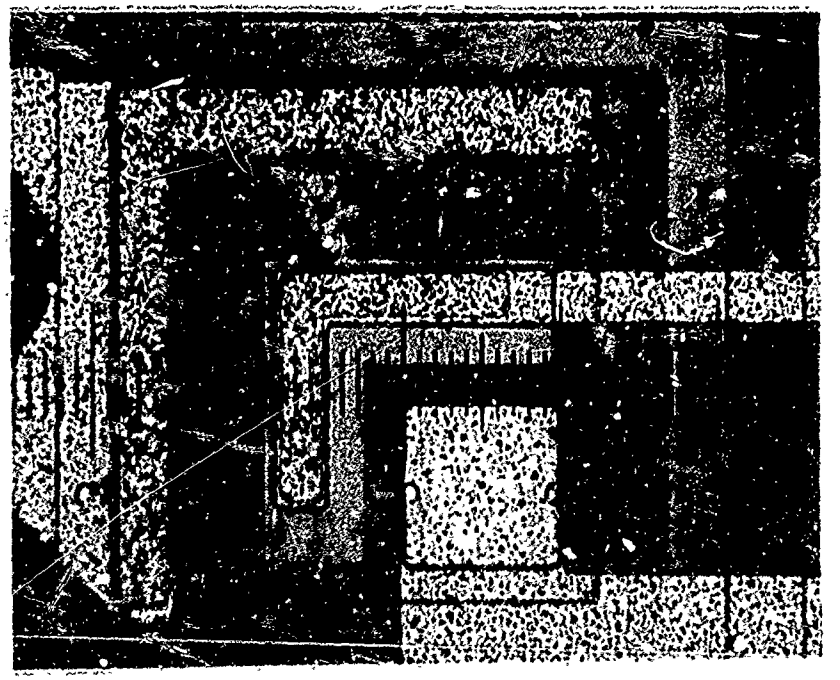


Fig. 35. Part B, special test circuit. (a) Overall view of chip. (b) View of transistors.

(c)



(d)

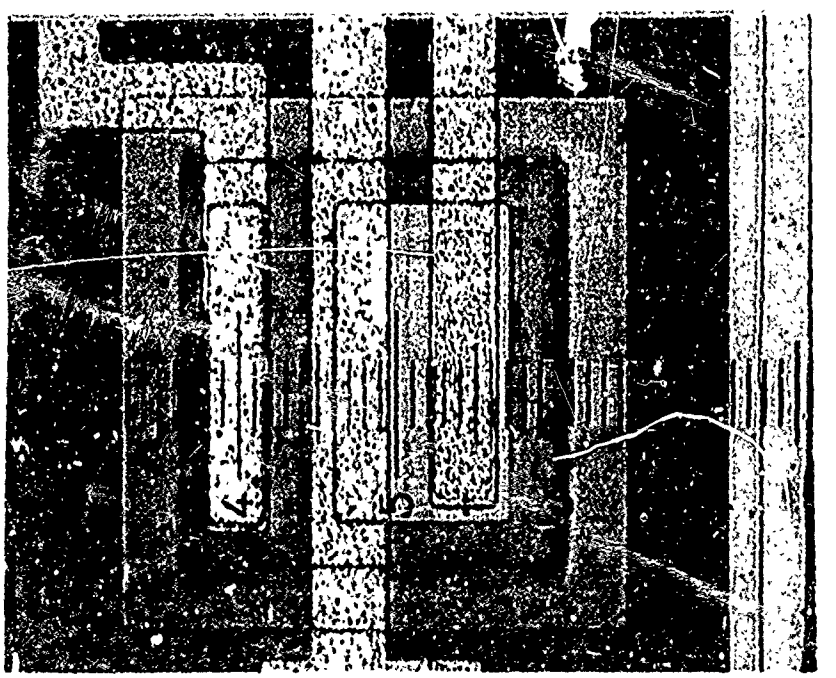


Fig. 35 (cont.). (c) Single transistor. Each small division is 6.25 μm. (d) Close-up of diode.

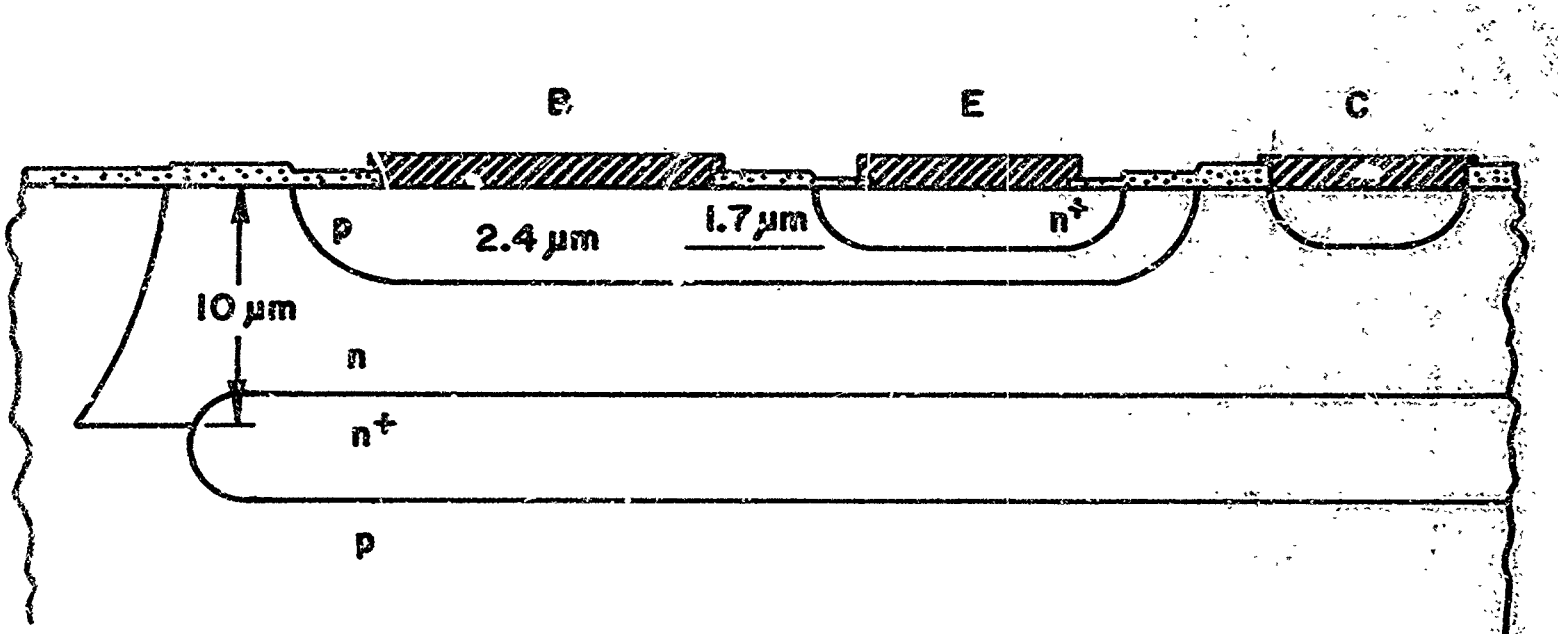


Fig. 36. Cross-section view of Part B. Dimensions are not to scale.

The base diffusion is 2.4 μm deep, with a sheet resistivity of 170 ohms per square. The emitter diffusion is 1.7 μm deep. The buried layer was not clearly revealed by lapping and etching but its presence is evident from the photos in Fig. 35. The oxide thicknesses were determined to be

collector	- 8,500 Å
isolation regions	- 6,800 Å
base	- 4,900 Å
emitter	- 1,500 Å

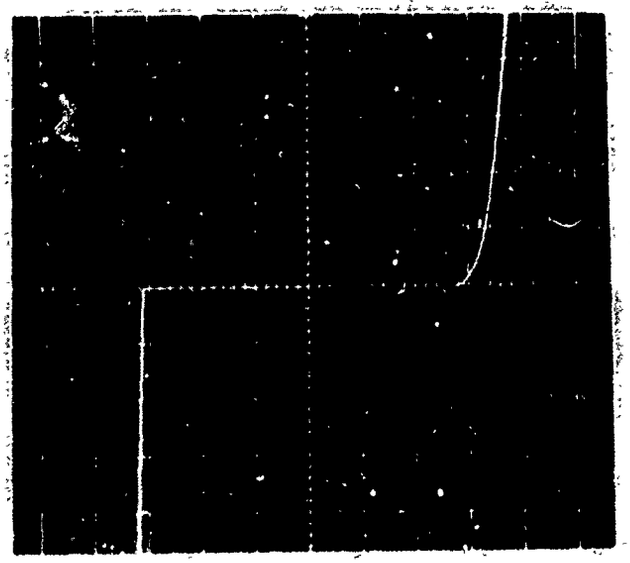
The manufacturer's specifications for the circuit are

h_{FE}	20-50	
BV_{CBO}	38-45	volts
BV_{CEO}	13-15	volts
BV_{EBO}	6.3-6.7	volts

Electrical characteristics are shown in Fig. 37. Part (a) is the diode characteristic. Part (b) is the emitter-base characteristic of the three transistors superimposed. In the forward direction two of the transistors have identical characteristics. The third transistor (pins 1 and 5) has an increased voltage drop due to the increased base width. In the reverse direction the shunting effect of the 1K resistor is apparent. Again two of the transistors have identical traces while the wide base device has an additional voltage drop.

Part (c) of the figure shows the collector-base characteristics. The switchback in the reverse bias direction is due to the emitter being connected to the base. Part (d) is the collector-emitter characteristic. The forward direction in the figure is under normal biasing, that is, the

(a)



(b)

Fig. 37. Electrical characteristics, Part B. (a) Diode characteristic. Forward, 0.5 ma/cm vertical and 0.2 v/cm horizontal. Reverse, 0.5 ma/cm, 2 v/cm. (b) Emitter-base junction characteristics. Forward, 10 ma/cm vertical, 0.2 v/cm horizontal. Reverse, 10 ma/cm, 2 v/cm. See text for explanation of multiple traces.

(c)

(d)

Fig. 37 (cont.). (c) Collector-base junction characteristics. Forward, 2 ma/cm, 0.2 v/cm. Reverse, 5 ma/cm, 10 v/cm. (d) Collector-emitter characteristics. Forward, 5 ma/cm, 10 v/cm, collector positive with respect to emitter. Reverse, 5 ma/cm, 1 v/cm, inverted polarity.

collector is positive with respect to the emitter. The switchback is more complex than usual, again due to the connection between emitter and base. In the reverse direction the collector is negative with respect to the emitter. The collector-base junction is then forward biased, and the 1K resistor appears in series with this junction. When the voltage is large enough to cause emitter-base breakdown the current begins to increase, with a switchback occurring due to carrier injection from the collector junction.

The emitter-base resistor was disconnected from one transistor by scratching open a portion of the metallization. The transistor characteristics were then found to be normal, with parameters very close to those specified by the manufacturer.

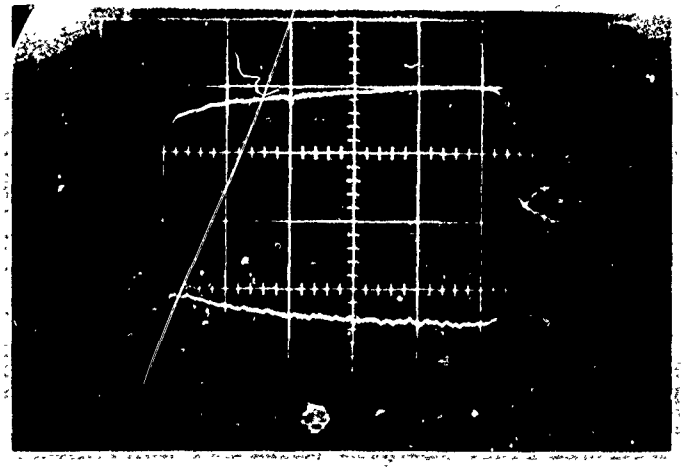
Second breakdown data on Device B will be presented in the following order. First the diode results will be given. This will be followed by the transistor emitter-base data, then the collector-base results. Finally the collector-emitter results will conclude this portion.

a) Part B - Diode Results

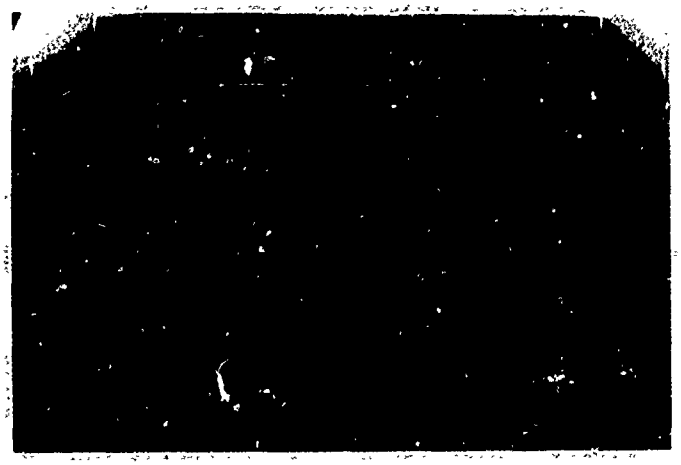
These diodes differed from those of Device A mainly by having a significantly larger area. The linear dimensions were very nearly 2½ times those of Device A (except for diffusion depths, which were nearly identical). The threshold power was consequently much greater. The base sheet resistivity was 170 ohms per square, compared to 130 ohms per square.

About 30 diodes were tested - 8 forward biased, 2 reverse biased at 77°K, and the rest reverse biased at room temperature. Representative waveforms of the latter tests are shown in Fig. 38, parts (a) and (b). The waveforms are similar to those of the emitter-base junctions of Device A. There are no "glitches" in Fig. 38, however, nor did they ever occur in these

(a)



(b)



(c)

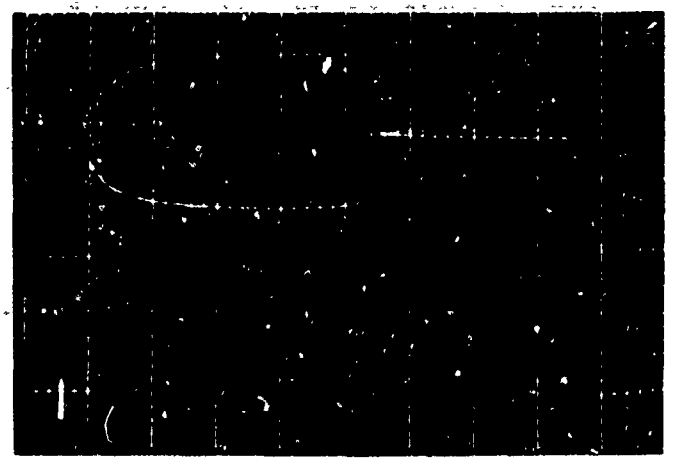


Fig. 38. Reverse bias waveforms for the diode of Part B. (a) Upper trace 20 v/cm, lower trace 500 ma/cm, time scale 200 nsec/cm. (b) 10 v/cm, 100 ma/cm, 100 μsec/cm. (c) 10 v/cm, 100 ma/cm, 100 μsec/cm. A 68 ohm resistor was connected in series and no damage resulted.

diodes. In most cases there was a gradual reversal of slope prior to second breakdown. The transition to second breakdown was usually, but not always, abrupt. In one case the slope change and second breakdown drop in voltage was so gradual that the waveform was more like that of a forward biased junction. In another case degradation of the reverse characteristic was noted even though no second breakdown transition was observed. In four cases second breakdown was not accompanied by observable damage, either electrical or visual. Figure 38(c) shows one such case. Note that the drop in voltage is significantly less than in the other two photographs, about 30% of the maximum voltage as opposed to 60%. In all such events the pulse width was relatively long. A 68 ohm resistor was connected in series with the diode in 3 of the 4 cases in order to obtain finer adjustment of the pulse amplitude.

In spite of the diversity in behavior the threshold power points of all diodes formed a smooth curve over nearly the whole range. Only at short pulse widths was there significant scatter of the data points.

Forward bias waveforms were similar to those of Device A. In general, there was a gradual transition to second breakdown. A drop in voltage did not necessarily signal the advent of permanent damage. Some of the waveforms are given in Fig. 39. Note in particular the waveforms in part (b). Only the third, highest amplitude pulse produced permanent damage. The waveform in (c) shows a small but noticeable drop after the large change in voltage. In a later part this will be shown to be related to damage progressing from the junction through the material on the low resistivity side to the contact.

Power versus pulse duration curves are drawn in Fig. 40. The

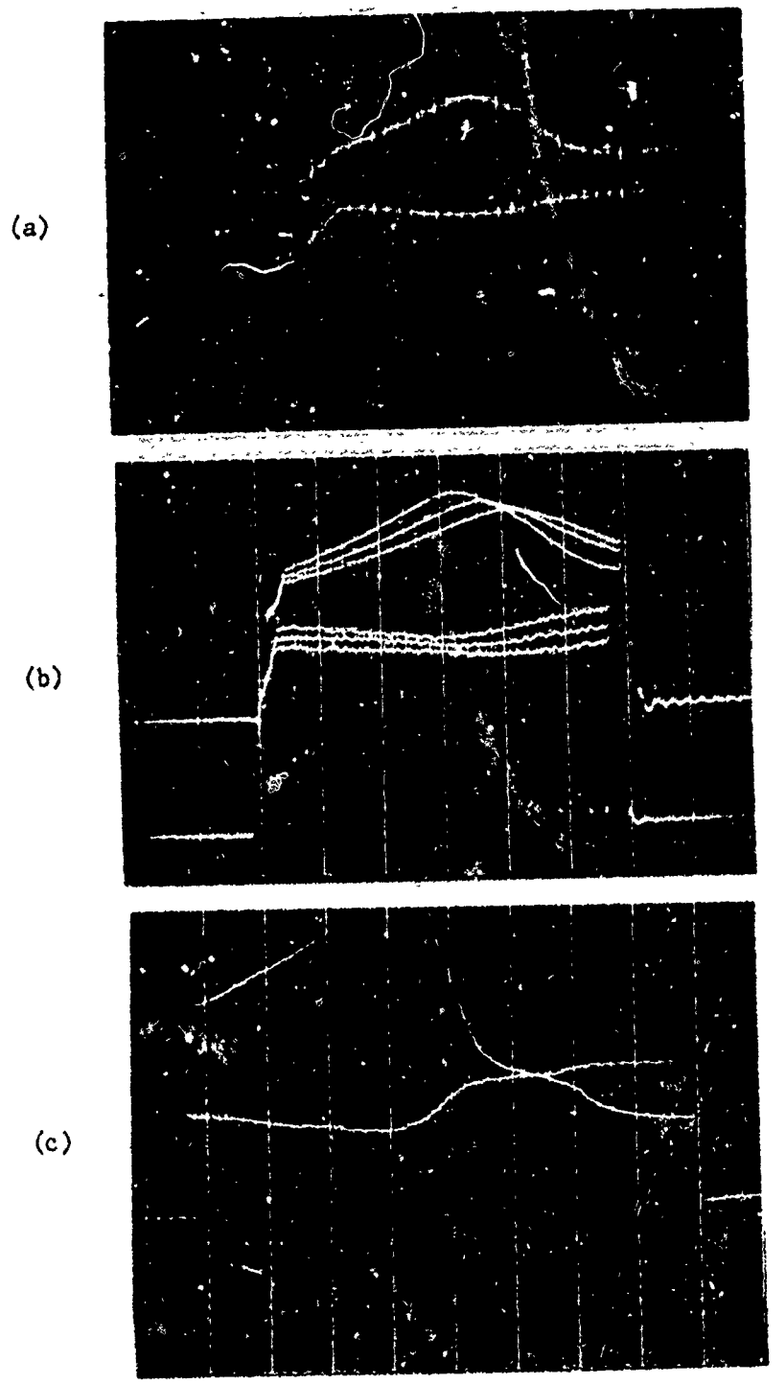


Fig. 39. Forward bias waveforms for the diode of Part B. (a) 20 v/cm, 2 A/cm, 200 nsec/cm. (b) 10 v/cm, 1 A/cm, 500 nsec/cm. (c) 5 v/cm, 500 ma/cm, 2 μsec/cm.

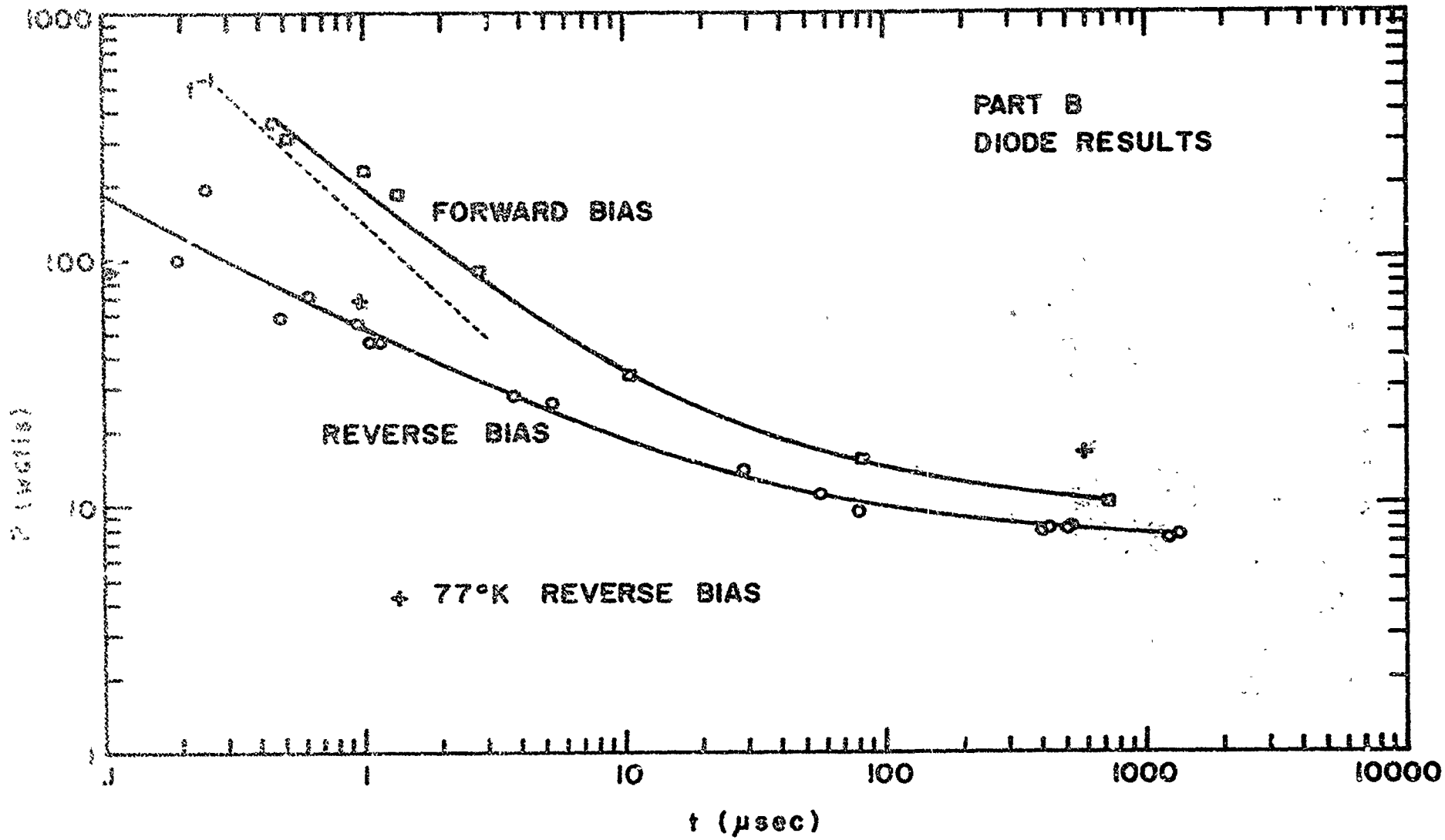


Fig. 40. Threshold power vs time curves for the diode of Part B.

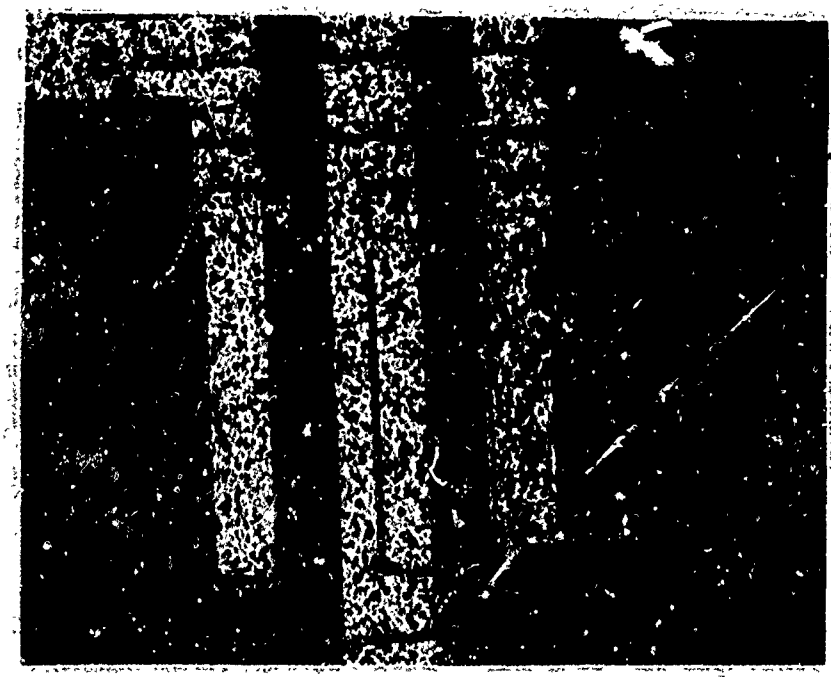
special features of the curves are made apparent by comparing with Fig. 27. The most obvious difference is the large increase in threshold power. This is due to the much larger size. Another difference is in the more gradual transition from the t^{-1} asymptote to the t^{0} asymptote. This again is a function of the size. Another interesting feature is that the threshold power is proportional to device area at short pulse widths for both forward and reverse biased junctions, while at long pulse times the power is proportional to linear dimensions.

This can be explained as follows. At short pulse widths the diode current is high. About half the power is dissipated at the junction and the rest in the base region between the active portion of the emitter junction and the base contact (this region has a resistance of about 35 ohms). At large pulse widths on the other hand, nearly all of the power is dissipated in the thin depletion layer of the junction. This contrasts with the SOS diodes of Sunshine where nearly all of the excess voltage above the breakdown voltage appeared across the space charge resistance of the junction.

Visual damage was apparent in most cases, with the exceptions already noted. Damage usually occurred in the base region but not in the emitter region. Damage often occurred to the base metallization and occasionally the aluminum stripe over the junction was affected. On rare occasions a short appeared between the stripe and one of the diode terminals.

Two damaged junctions are shown in Fig. 41. Part (a) shows typical damage from a reverse bias test. Damage extends from the base contact to the edge of the aluminum stripe. When the aluminum is removed in such cases the track can be seen to extend to the junction. In part (b) the stripe has melted. The base metallization also shows damage. There is no damage in the

(a)



(b)

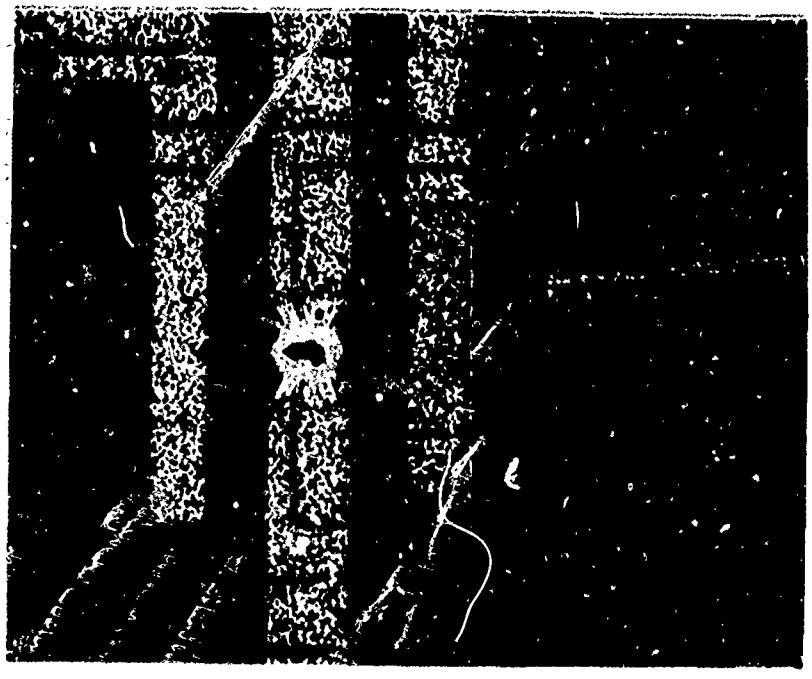


Fig. 41. Diode damage due to reverse bias pulses.

emitter region. The pulse waveforms for this particular diode are those shown in Fig. 38 part (b). The pulse width up to second breakdown is 520 μsec and the average pulse power is 8.4 watts. After second breakdown has occurred the diode dissipates about 6.7 watts for 380 μsec. It is difficult to ascertain exactly when the aluminum melted. Clearly the damage must have occurred after filament formation. However, it cannot be ruled out that melting may have occurred after the pulse terminated. Unfortunately, the one-dimensional heat flow calculations of Section II provide little insight for events with such long pulses.

To summarize the results of the diode data for Device B, we can state that the differences in behavior from Device A can be attributed to the difference in size. Anomalous behavior of a few devices was observed.

b) Part B - Emitter-Base Junctions

The emitter-base junctions were stressed only in the reverse direction and only at room temperature. Differences in junction characteristics, breakdown waveforms, and damage can be correlated with the differences in geometry of the transistors. As would be expected, there is a close relationship between the second breakdown characteristics of the emitter-base junctions and the diodes previously discussed.

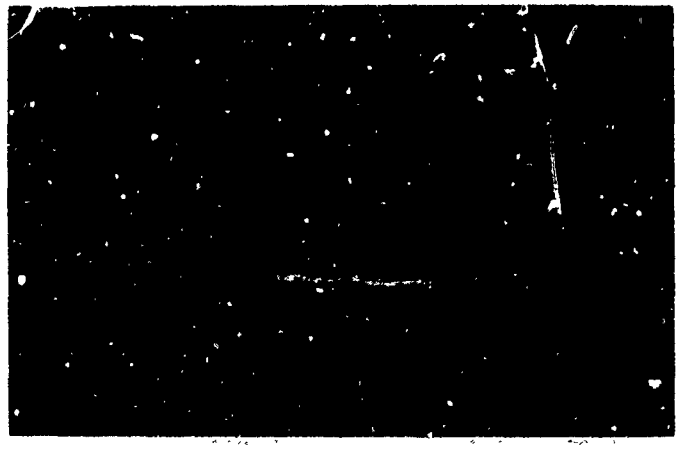
Current-voltage waveforms fell into distinct categories, some occurring more commonly than others. The most usual waveform showed a slight rise in voltage to a broad maximum, a slight decrease, and then an abrupt drop to a low sustaining level. In one variation the transition was first to an intermediate level for a short period. This happened only on rare occasions and always at short pulse widths. Degradation of the junction characteristic did not always occur in such cases. In a second variation a

small drop in voltage (about 2 volts) took place after the main transition. This happened only with longer pulses.

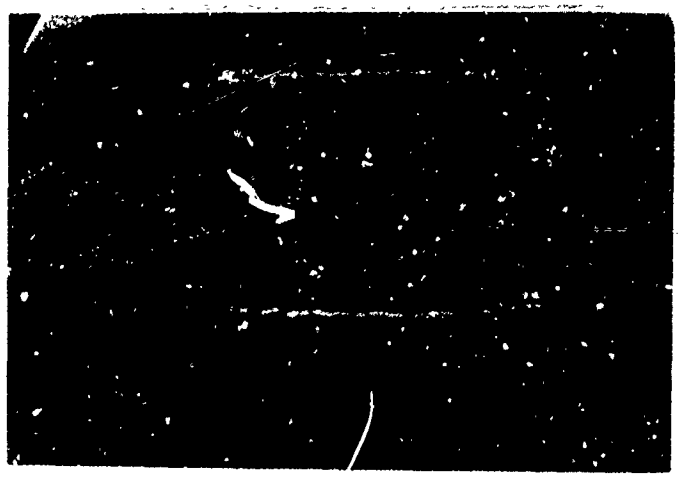
Waveforms are shown in Fig. 42. Parts (a) through (d) of the figure are devices on the same chip. Part (a) is the diode. Part (b) is the "normal" transistor, pins 1-13. It differs from the diode waveforms only in having a lower voltage. The differences at the leading edge are due to the pulse generator and can be ignored. Part (c) of the figure is the waveform of transistor 1-3, the "wide-emitter" transistor. The voltage and current are slightly reduced because the pulse generator output was reduced. The time delay before second breakdown is correspondingly increased, but otherwise the waveforms are the same as (b). Part (d) of the figure is the "wide-base" transistor, pins 1-5. This is an example of the variation with the transition to an intermediate level. The junction was damaged in this case. One is tempted to postulate that the first transition corresponds to a filament forming in the base region, followed by elongation to the junction. The evidence is too sparse to confirm this contention, however. Furthermore, Budenstein et al., found this sequence of events occurring only with long pulses in their diodes.

Figure 42 (e) and (f) show breakdown waveforms with longer pulses. The waveforms in (e) are between pins 1 and 3, while those in (f) are between pins 1 and 5. Note the small drop in voltage after second breakdown has occurred in (f). This is an illustration of the second variation in waveform. Examination under a microscope revealed a stress mark extending all the way from the base contact to the emitter contact of the transistor in (f), but only from the base contact to the junction in (e). The curve tracer showed emitter-base damage in both transistors and collector-emitter degradation of

(a)



(b)



(c)

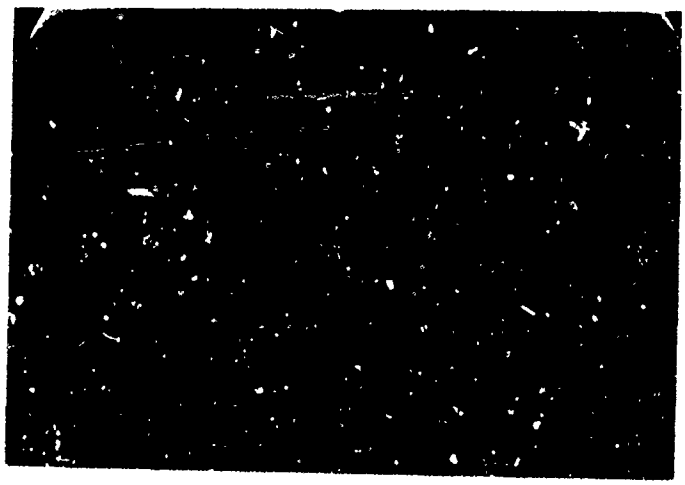


Fig. 42. Emitter-base pulse waveforms for Part B, reverse bias. (a) 20 v/cm, 500 ma/cm, 100 nsec/cm. (b) 10 v/cm, 500 ma/cm, 100 nsec/cm. (c) 10 v/cm, 500 ma/cm, 100 nsec/cm.

(d)

(e)

(f)

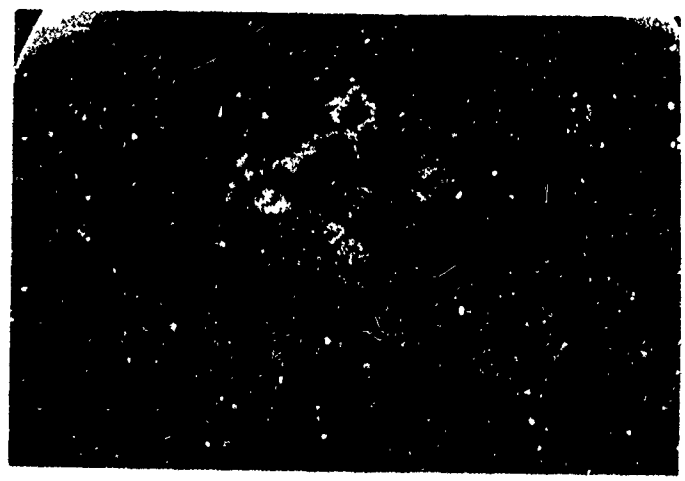


Fig. 42 (cont.). (d) 20 v/cm, 500 ma/cm, 100 nsec/cm. (e) 10 v/cm, 200 ma/cm, 1 μsec/cm. (f) 20 v/cm, 200 ma/cm, 1 μsec/cm.

the transistor in (f). Normally, second breakdown involves damage to the junction and the base region only. If the duration in second breakdown is long enough, damage extends all the way through the emitter to the emitter window, accompanied by a further drop in voltage. Nearly always when this variation was observed the distinction in visible and electrical damage could be made.

The waveforms show some correlation with the spacing from junction to emitter contact on each of the transistors. The spacing was largest between pins 1 and 3, and this transistor consistently showed the largest sustaining voltage after second breakdown. It was higher by at least 2 volts. This transistor was also the least likely to show a second transition after second breakdown. The explanation is straightforward. Once second breakdown has occurred and a melt column bridges the base region, the emitter material serves as a ballast resistor. The greater the spacing between the junction and the emitter contact, the greater the resistance and the greater the sustaining voltage. If the distance is short and the pulse length is long the filament is likely to bridge the emitter region, causing the three observed effects: a drop in voltage, a stress mark bridging the contacts, and changes in the collector-emitter breakdown characteristic.

The effects of increased base width are evident prior to second breakdown as an increase in voltage required to induce second breakdown with a given pulse width. Figure 42 again illustrates this when part (d) is compared with either (c) or (b). The current is nearly the same, but the voltage in (d) is about 50% greater.

The diode results verify these observations. The diode has about the same amount of active emitter periphery but larger emitter and base

widths from junction to contact. We would therefore expect a higher voltage before second breakdown and a higher sustaining voltage after second breakdown. Both these conclusions are substantiated in Fig. 42(a).

In Fig. 43 the threshold power versus pulse duration curves are plotted. The upper curve is for the wide-base transistor and makes clear the greater power requirements, especially at short pulse widths. Data points for the other two transistors are plotted separately but there is no clear effect due to the differences in emitter width and only one curve is drawn. Both curves are substantially lower than those of the diode, Fig. 40, where the base width was considerably greater. Recall also that the junction to base contact width had a pronounced effect on threshold power for Part A (Fig. 27).

A photomicrograph of a typical emitter-base zap is shown in Fig. 44(a). Junction damage is clearly seen, with the track extending through the base to the base contact. This type of damage was observed in about 75% of the cases. The breakdown waveform for this junction is that shown in Fig. 42(e). Part (b) of Fig. 44 shows the damage produced by the pulse of Fig. 42(f). Damage extends all the way from contact to contact.

It is clear that the emitter-base pulsing has yielded a richer variety of results than the diode tests. In the first place the important role of the base material in increasing the power handling capability has been corroborated. The effect of the emitter material has also been noted. Apparently the only appreciable modification is on post-breakdown behavior, where a higher sustaining voltage results. More importantly, increased emitter width has the desirable effect of decreasing the transistor's susceptibility to emitter to base shorts during second breakdown.

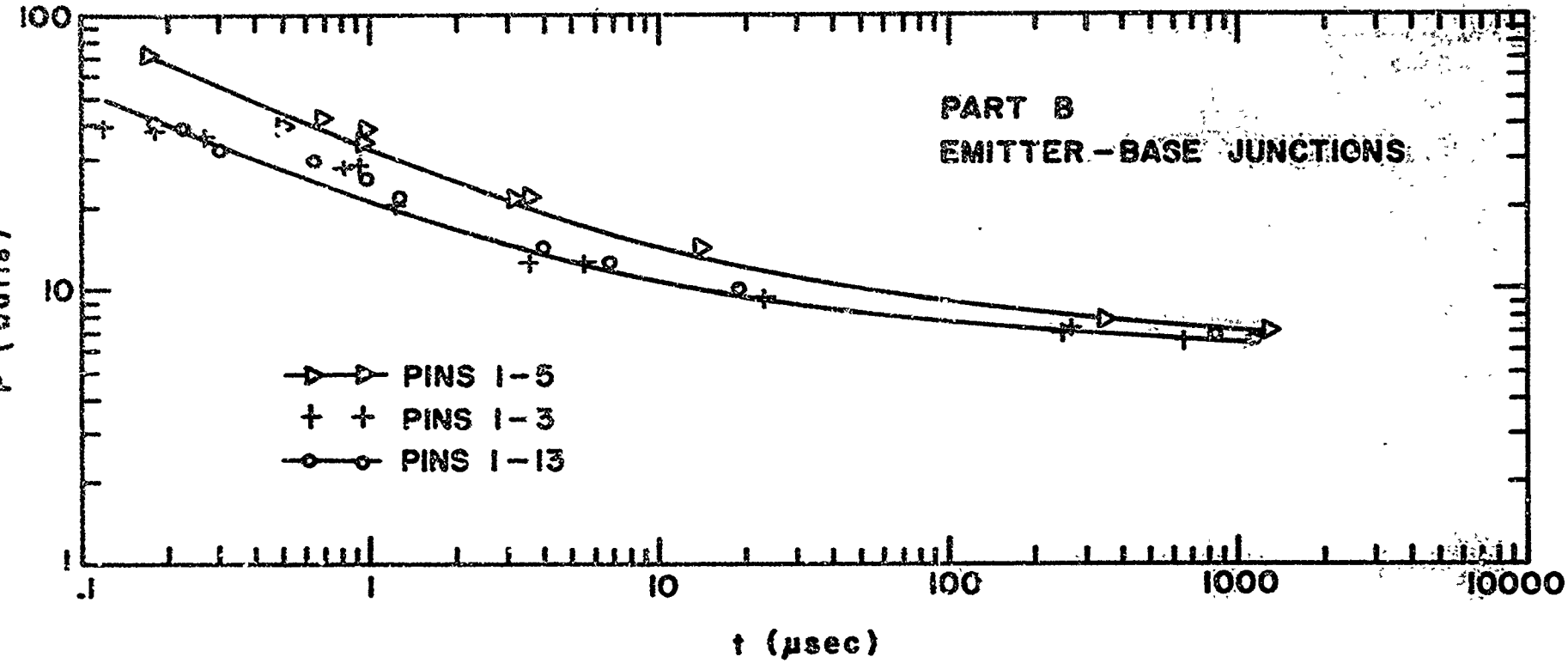
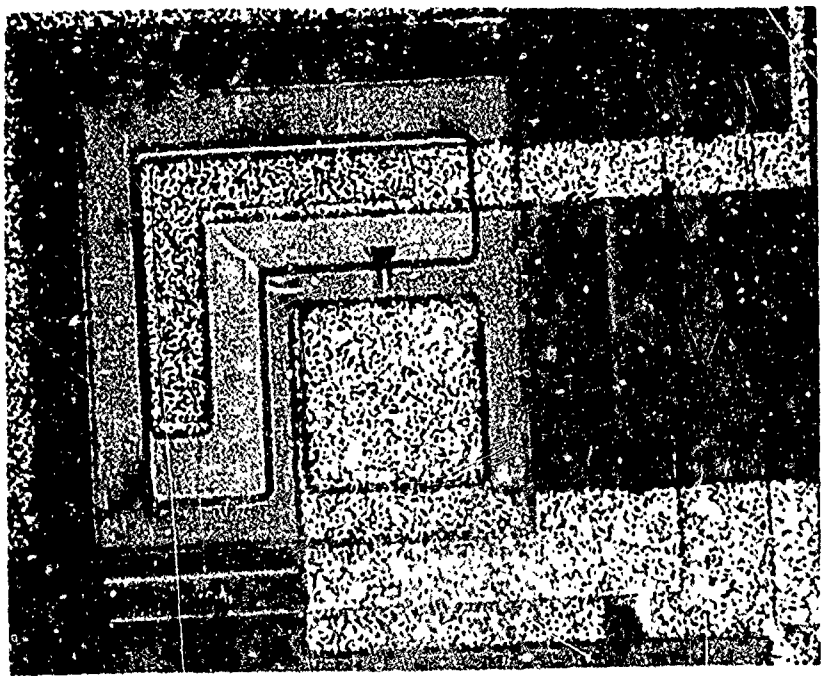


Fig. 43. Power versus delay time curves for Part B emitter-base junctions. The upper curve is for the wide-base transistor.

(a)



(b)

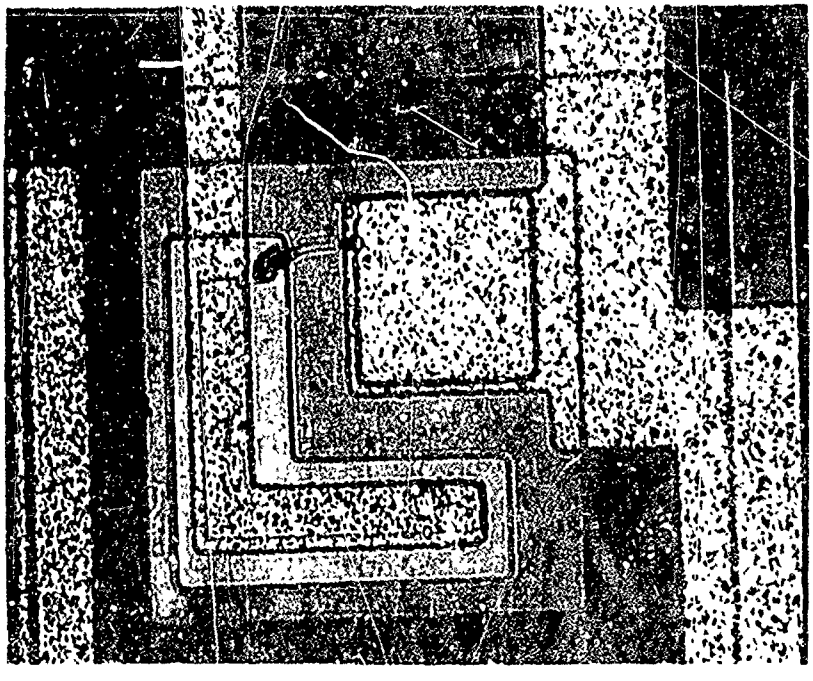


Fig. 44. Emitter-base junction damage. (a) Damage produced by the waveform of Fig. 42(e). (b) Damage from the pulse in Fig. 42(f).

Unfortunately, the changes in transistor beta caused by emitter-base second breakdown were not investigated in this study. One is led to speculate that the effects would be minimal since the junction forward characteristics are not affected and because the damage occurs at the surface. Others have found major changes in beta;[30,31] however, the tests were performed on discrete devices. Power levels were higher and damage tended to be more massive.

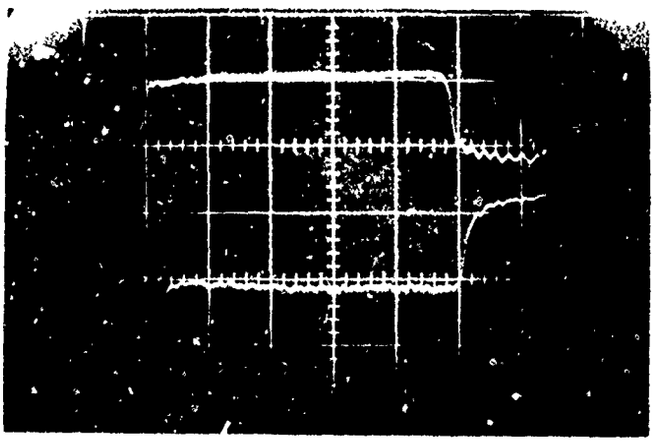
There is little change in the collector-emitter breakdown characteristics of either polarity as long as damage is confined to the emitter junction and the base region. If the damage bridges the emitter-base contacts noticeable changes are introduced especially in the inverted direction with the collector negatively biased with respect to the base.

c) Part B - Collector-Base Junctions

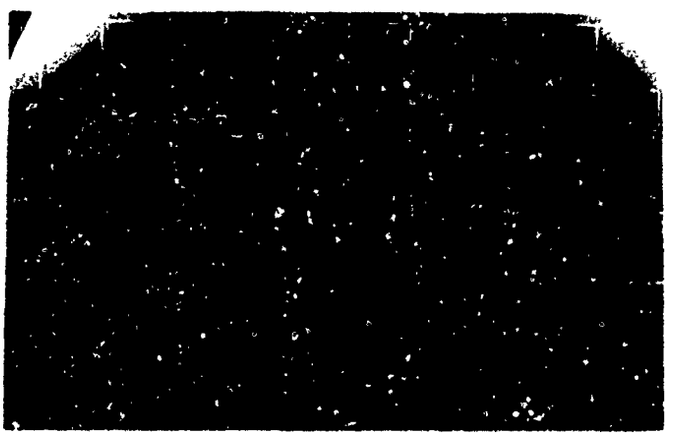
Collector-base junctions were tested on 14 parts, 10 in the reverse bias direction, 2 in the forward direction, and 2 in the reverse bias direction at liquid nitrogen temperature. Good, consistent data were obtained as far as pulse waveform and threshold power were concerned. Device damage showed some variability. As in previous tests the occurrence of second breakdown resulted in degradation of device characteristics in nearly every instance. Unlike previous tests the base contact to junction distance had only a barely discernable effect on threshold energies.

The three transistors on each chip had virtually identical waveforms for the same pulse conditions. Representative waveforms are shown in Fig. 45. (a), (b), and (c) are reverse bias pulses. For pulse durations greater than 1 μsec the terminal voltage remains at a high level for a fraction of a μsec, then drops to a slightly lower level as the device

(a)



(b)



(c)

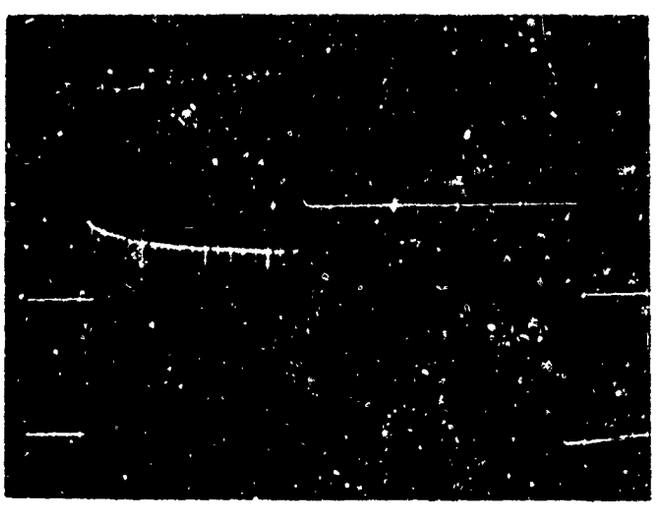


Fig. 45. Waveforms for collector-base junctions, Part B. (a) 20 v/cm, 500 ma/cm, 200 nsec/cm. (b) 20 v/cm, 200 ma/cm, 1 usec/cm. (c) 10 v/cm, 100 ma/cm, 100 μsec/cm.

(d)

Fig. 45 (cont.). (d) Forward bias mode, 5 v/cm, 500 ma/cm, 200 μsec/cm. Permanent damage occurs at the small drop in voltage.

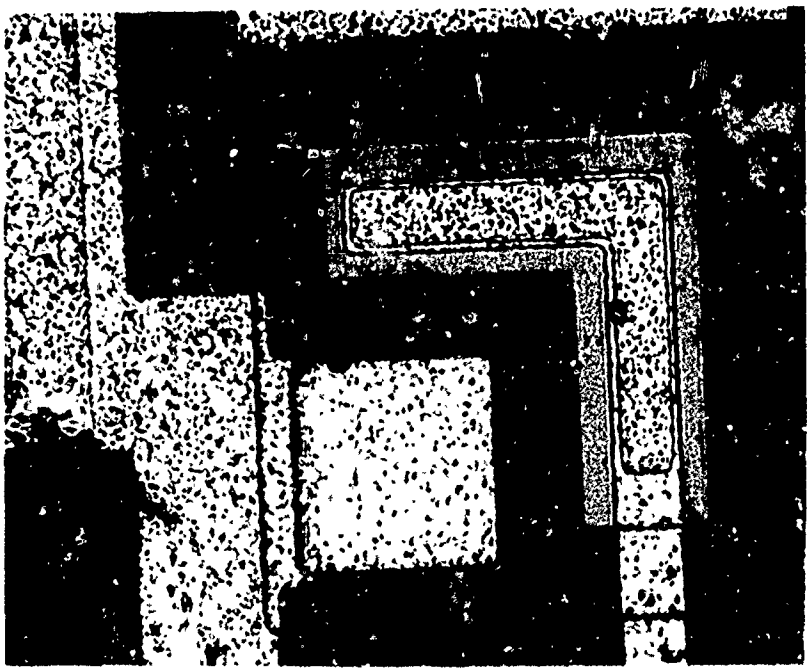
begins to oscillate. Whether or not oscillation takes place, the transistor is damaged only when there is a pronounced drop in voltage to about the 50% level.

With short forward biased pul .s the voltage waveform was much like that of earlier forward biased junctions. The voltage increased to a broad maximum and then began to decrease. Transistors had to be checked on a curve tracer after each pulse to ascertain when damage occurred. Figure 45 (d) shows a forward bias pulse for a long pulse width. The waveform is different from previous cases. The broad maximum is still present, but the second breakdown transition is unusually small. It must be labelled as second breakdown, however, since damage occurred in all transistors tested at long pulse widths if and only if the transition was present.

Visible damage fell into two categories. Illustrations appear in Fig. 46. Part (a) shows the damage produced by the pulse in Fig. 45 (b). For pulse widths less than 10 μsec the damage was of this type and looked like the damage seen in a forward or reverse biased emitter-base junction. Figure 46 (b) shows the other type of damage commonly seen with long pulses. A corner of the base metallization shows clear evidence of heating. The corner of the emitter metallization also shows signs of reaching high temperatures. When this type of damage occurred subsequent pulses increased the amount of damage but no stress marks of the type in (a) appeared.

Electrical damage followed from second breakdown in all but 4 cases. All of these were for long pulse widths and 3 had a 68 ohm resistor connected in series with the device. Damage consisted of changes in the reverse bias breakdown characteristics of the emitter junction, the collector junction, or from collector to emitter, either polarity. There was a correlation between

(a)



(b)

Fig. 46. Damage due to collector-base pulses. (a) Damage at short pulses. (b) Contact damage at long pulse widths.

electrical and visual damage. When the stress mark appeared at the surface, as in Fig. 46 (a), the emitter junction was always degraded and sometimes the collector-emitter breakdown in the inverted direction was changed. Surprisingly enough the collector-base junction characteristics were often unchanged. Whenever the damage occurred only to the base metallization the collector junction showed deterioration as well as the collector-emitter breakdown in the normal direction. Often visible damage of either type was accompanied by damage to all junctions.

The evidence indicates the presence of two coexisting current carrying paths. One is from the collector contact down to the buried layer, along the buried subcollector, back up through the epi layer, across the avalanching collector junction, and to the base contact. The other path is along the surface from the collector contact to the reverse biased collector junction, through the base region, into the emitter, across the reverse biased emitter junction, and through the base region to the base contact. At any given terminal voltage, each path will conduct different amounts of current. The heat produced along each path will depend on the current density and the electric field along each portion of the path. However, the paths are close enough to each other that the thermal responses overlap. (The diffusion length for heat flow for a 1 μsec pulse is about 10 μm, the depth of the buried layer.) Apparently there is a considerable amount of heat produced near the base contact due to current flow along each path. Whether a filament forms first at the collector junction just beneath the base contact or at the surface near the emitter junction depends on the temperature distribution. This, in turn, depends on the current division between the two paths for a particular applied voltage. At high pulse voltages apparently

the conditions for filament formation are reached first between the base contact and the emitter. Wherever the breakdown occurs, the two junctions are close enough near the base contact that when one is damaged the melted region is very likely to damage the other also.

Brown, Holder, and Ruwe[30] describe an anomalous mode of failure in discrete transistors in terms of a surface avalanche which they refer to as "Holder avalanche." They also observed emitter junction damage, both electrical and visible, when pulsing between collector and base. Their explanation is along the lines of that given here even though the geometry of a discrete transistor is somewhat different.

The threshold damage curve is plotted in Fig. 47. The curve is drawn for the data points corresponding to the "normal" and "wide-emitter" transistor geometries, which appear to have no distinct differences. A separate curve is not drawn for the "wide-base" geometry but the data points are seen to lie slightly above the previous curve over the entire range. This again points to the existence of a path that carries a portion of the current parallel to the surface at the base contact; otherwise, the increased spacing would have no effect.

The curve shows nearly a t^{-1} asymptote at short pulses. This implies heat generation over a fairly large volume - despite the large drop at the junction - such that the heat conducted away is only a small portion of the total heat produced.

The power level at 1 msec is the same as the emitter-base junctions, Fig. 43. At ½ μsec the power is greater by a factor of about 2. A rather surprising result is found when the data of Fig. 47 is compared to that of the diode, Fig. 40. The results almost coincide for reverse, forward, and low temperature cases.

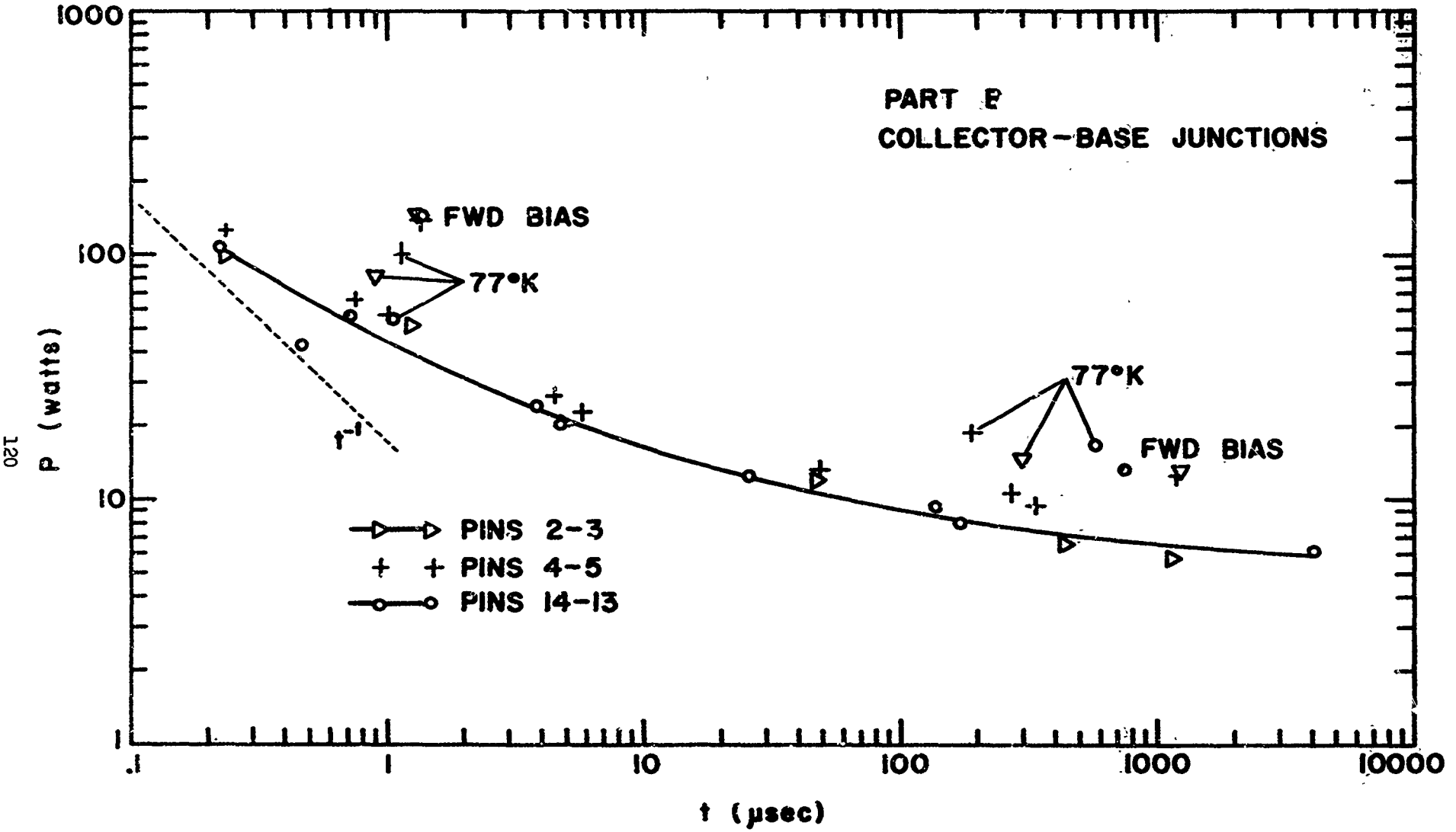


Fig. 47. Threshold damage for collector-base junctions, Device B. Data points for the wide-base transistor are consistently higher by a small amount in the reverse biased mode.

d) Part B - Collector-emitter Results

Collector-emitter pulsing would be expected to be more complex because of the interacting junctions, and indeed this proved to be the case. At first glance we would expect greater power handling capability because of the greater area between the emitter and collector contacts. Again this proved to be true. Pulse waveforms were more complicated than in previous test devices. There were two distinct voltage drops, both of which could produce contact damage but only one of which was associated with electrical damage. This will be explained in greater detail later on, with a breakdown model postulated to explain the results.

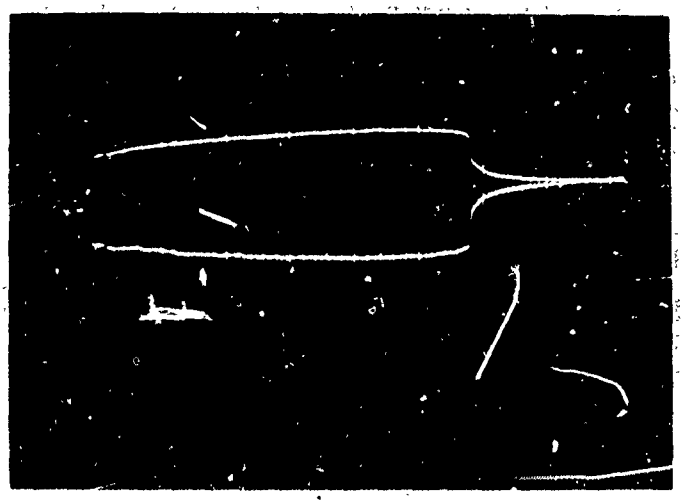
Another distinctive feature was the common occurrence of damage to metallization runs far from the junction. This is a consequence of the high currents encountered during breakdown.

Seven parts were tested. Some underwent the normal procedure while others were used in multiple or repetitive pulse testing in a closed-circuit TV system. All transistors were stressed with the collector positive with respect to the emitter and at room temperature only.

In the initial tests a significant drop in voltage was noted after a certain delay time which depended on pulse amplitude. This was interpreted as second breakdown. Figure 48 shows several waveforms. At low pulse amplitudes (parts (a) and (b) of the figure) the voltage change is significant and well defined. At high pulse amplitudes (parts (c) and (d) of the figure) although the voltage drop is the same in amplitude it is a much smaller percentage of the total voltage. Also a preceding, much smaller step in voltage can be seen.

This response was consistently observed in every transistor. It

(a)



(b)

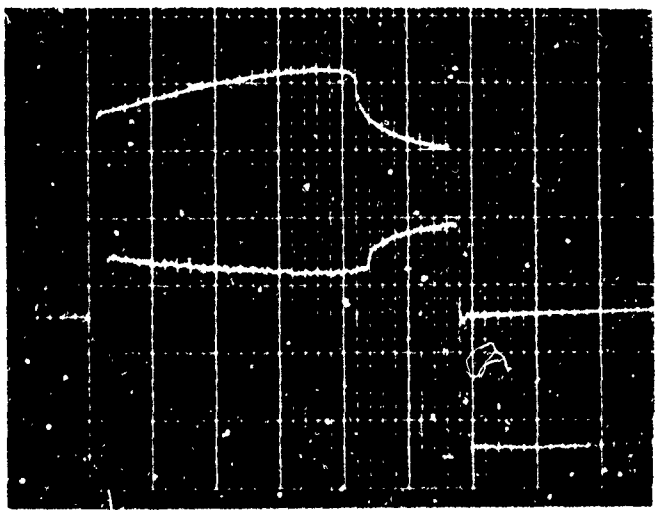
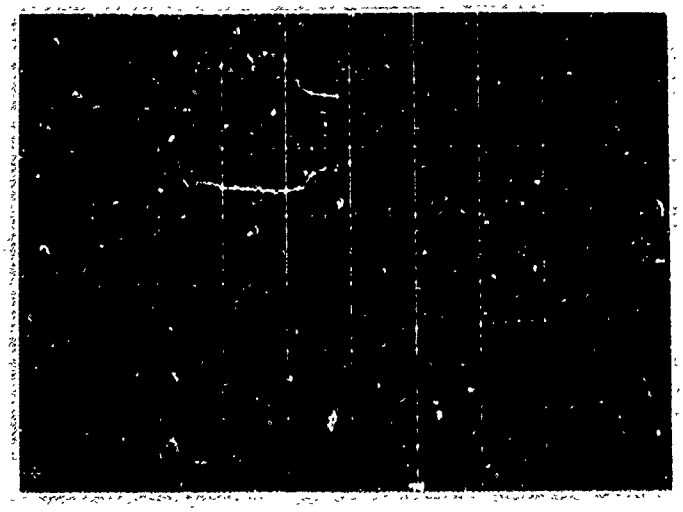


Fig. 48. Collector-emitter breakdown waveforms for transistors of Device B. (a) 10 v/cm, 200 ma/cm, 10 μsec/cm. (b) 10 v/cm, 500 ma/cm, 1 μsec/cm.

(c)



(d)

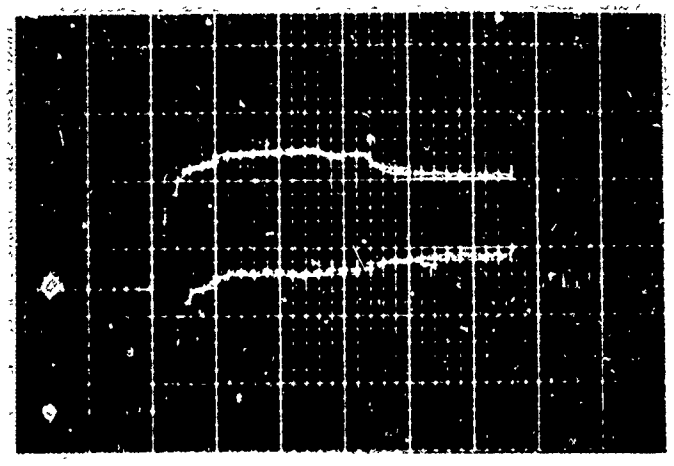


Fig. 48 (cont.). (c) 10 v/cm, 500 ma/cm, 500 nsec/cm. (d) 20 v/cm, 1 A/cm, 200 nsec/cm.

never resulted in electrical damage - all terminal pairs showed the same characteristics before and after such pulses. Because of this the transistor could be pulsed repeatedly into "second breakdown" over a wide range of pulse conditions and still yield exactly the same waveforms. No differences could be detected among the three types of transistors either in threshold energies or in waveforms.

The origin of the voltage discontinuity must be thermal in origin because the delay time depends on pulse amplitude. In the range from 3 μsec to 4 msec the threshold power versus delay time plot was a smooth curve slightly higher than the collector-base threshold curve, Figure 47, at long pulse widths but greater by a factor of 2-1/2 at short pulse widths. Below 1 μsec the curve seemed to level off in a constant power dependence. This is not certain, however, since the transition is poorly defined at low pulse widths and only a few data points were taken. If true, it would point to another mechanism becoming dominant such as current induced second breakdown.

Although electrical damage was not observed with this type of "second breakdown" contact damage was produced. The damage consisted of darkening of the emitter metallization over a sizeable portion of the contact, in contrast to the usual damage where the aluminum melted at a specific site. To determine the relationship between contact darkening and the change in voltage a video camera was mounted over a microscope and connected to a video tape recorder and TV monitor. Pulses were applied singly or at a low prf (about 10 pps) for tens of seconds at a time.

The conclusions were that the voltage drop was not absolutely necessary for contact damage but that damage was induced at a distinctly accelerated rate when the voltage change was present. When the pulse

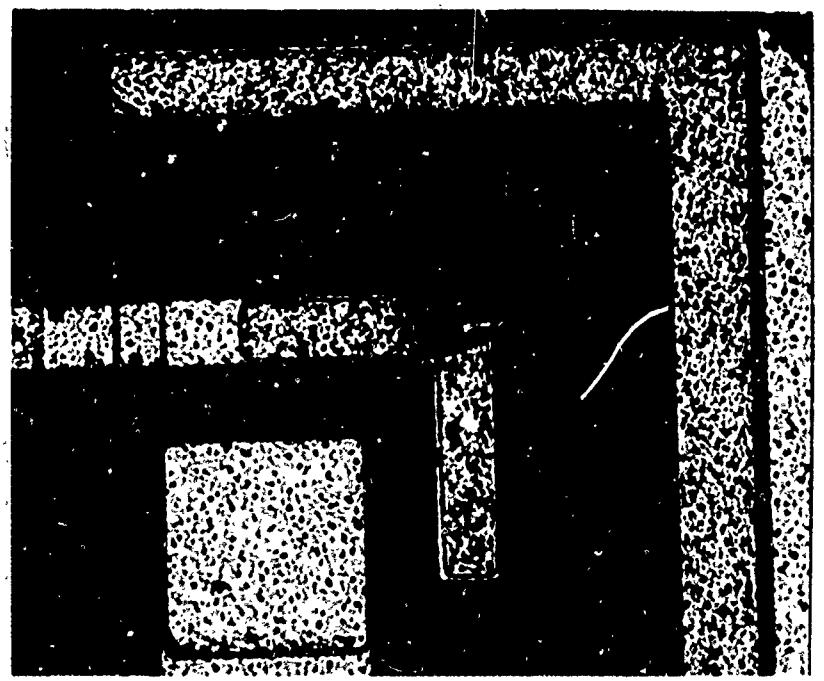
duration was less than the delay time, some damage was evident after a sufficient number of pulses had been applied. When the pulse duration was long enough to include the transition a single pulse was sufficient to produce noticeable damage.

Figure 49(a) shows emitter contact damage to a transistor after two pulses. "Second breakdown" occurred only during the second pulse. Unfortunately, the device was not opened for examination until after the second pulse so the extent of contact damage after the first pulse, if any, is not known. Notice that there is no distinct evidence of melting and that damage appears over a fairly large region. Figure 49(b) shows the appearance of another contact after 5-10 minutes of low prf pulsing at sub-breakdown pulse widths. Although the contact is heavily damaged, there is still no change in junction breakdown characteristics at this point. Degradation did not occur until further pulsing eventually caused the darkened metal to spread out and reach the emitter-base junction.

Because of the lack of junction degradation even after many excursions into what appeared to be second breakdown, and because the transition at high pulse amplitudes was not pronounced, it was decided to pulse the transistors at increasingly higher levels until permanent junction damage did occur. As the pulse energy was increased, either by increasing the pulse width or the pulse amplitude, a new breakdown phase came into existence, heralded by a further reduction in voltage. A single pulse incorporating this new feature was sufficient to produce dramatic effects. These included permanent device degradation as well as visible, often massive, contact damage.

Examples of pulse waveforms under these conditions are shown in

(a)



(b)

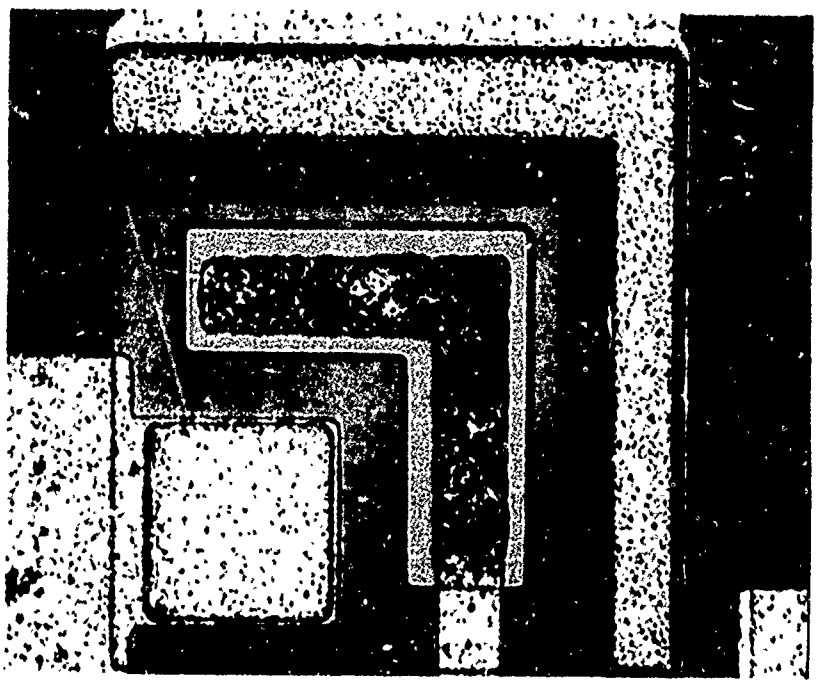


Fig. 49. Contact damage produced by pulses at/or just below first transition levels. No electrical degradation resulted. (a) After 2 pulses. (b) A different transistor after 5-10 minutes of low prf pulsing.

Figure 50. The first photo is for very long pulse lengths. The initial voltage peak and two transitions are completed in about 750 μsec. The residual voltage is only about 5 volts and has several minor fluctuations. The second photograph in the figure is at intermediate pulse lengths. Here the details are much clearer. The voltage initially rises from device heating to a maximum, then declines by 5 volts in the first transition. The voltage continues to decline slowly at this intermediate stage until final breakdown takes place at about 19 μsec. This final drop is comparatively gradual and occurs in two stages. The residual voltage is again about 5 volts.

The last waveform in the figure is representative of short pulses. The voltage overshoots at the beginning, then rises again due to heating. The first break occurs about 300 nsec after the beginning of the pulse. This is followed by a gradual decline until final breakdown takes place about 800 nsec from the start of the pulse. Just before the pulse terminates, the voltage rises sharply - this is associated with melting of the metallization run.

Several damaged transistors can be seen in the photomicrographs of Figure 51. The first photo shows a failure induced by a short pulse. Aluminum at the corner of the contact has melted and the stress mark extends into the emitter region. This is the only visible damage. There is no evidence of damage to the emitter junction, the collector junction, or the collector contact. A curve tracer did show junction damage, however.

The second photomicrograph is the same transistor shown in Fig. 49(a). Several more pulses have been applied, the last one sufficient to induce the second voltage transition.

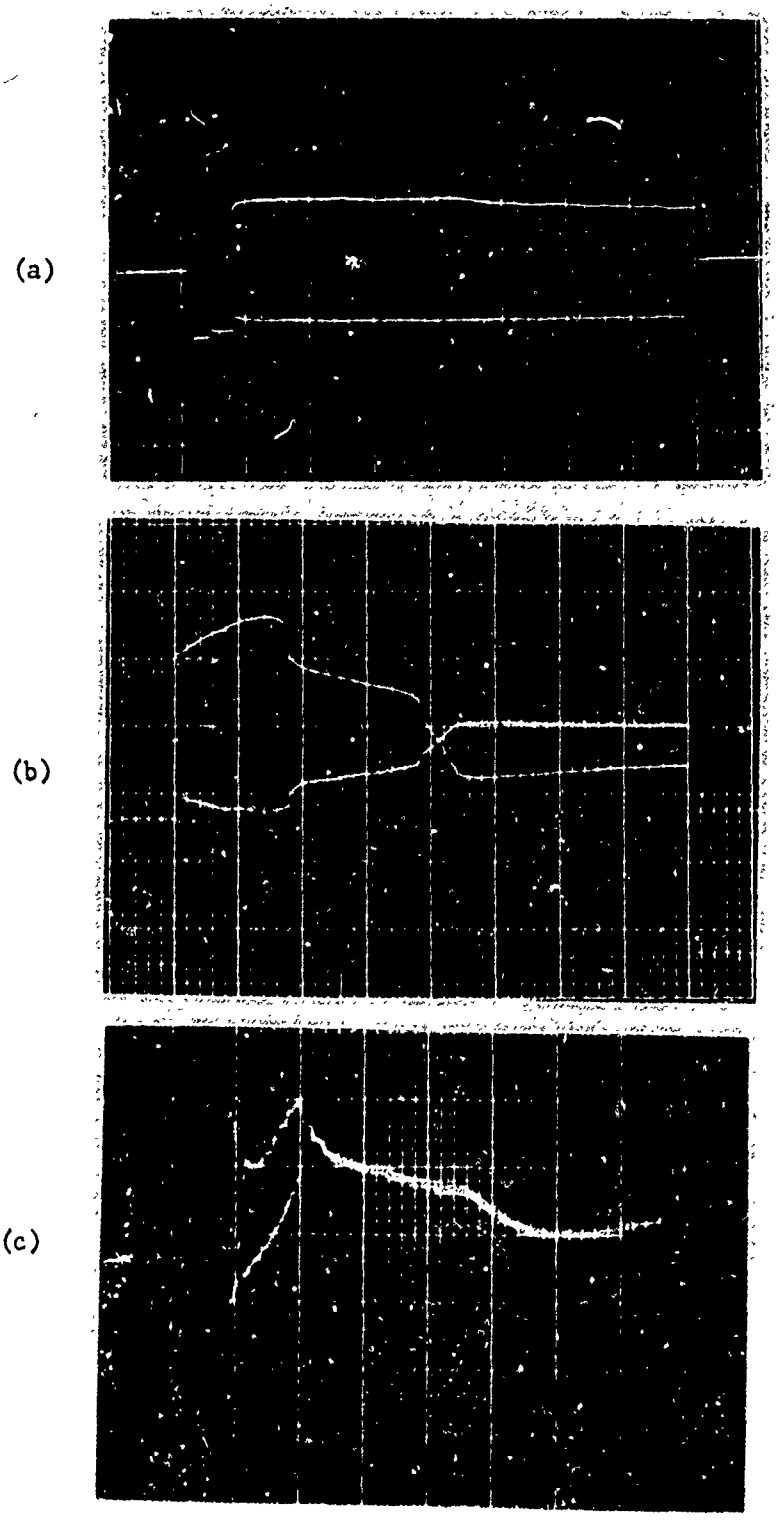
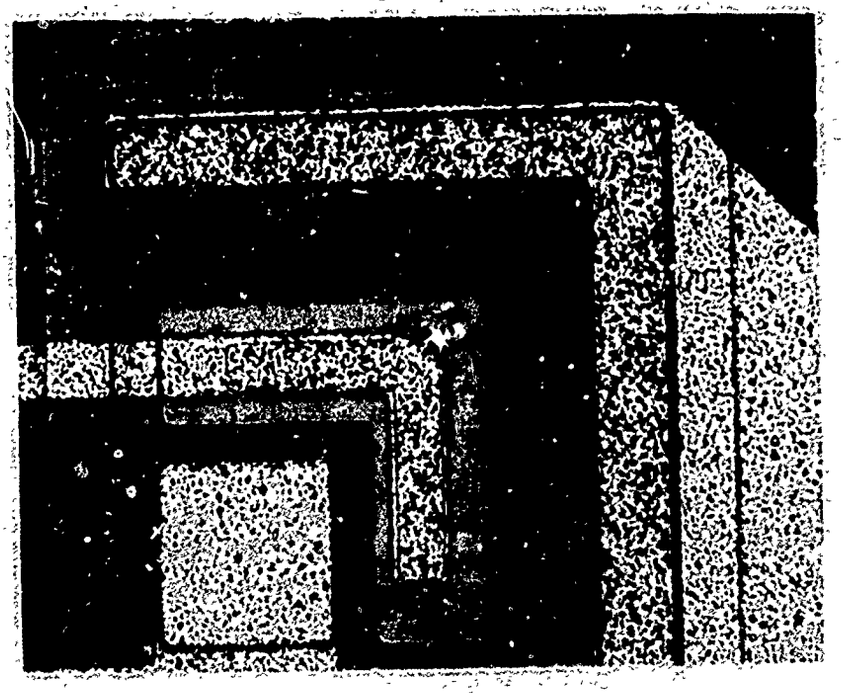


Fig. 50. Collector-emitter breakdown waveforms showing the second transition. (a) 5 v/cm, 500 ma/cm, 1 msec/cm. (b) 10 v/cm, 500 ma/cm, 5 μsec/cm. (c) 20 v/cm, 1 A/cm, 200 nsec/cm.

(a)



(b)

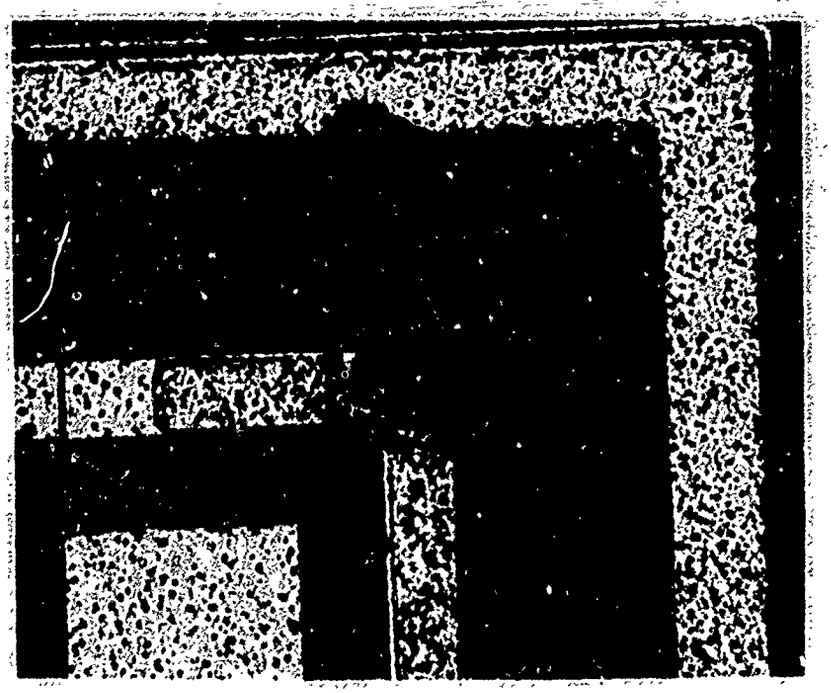
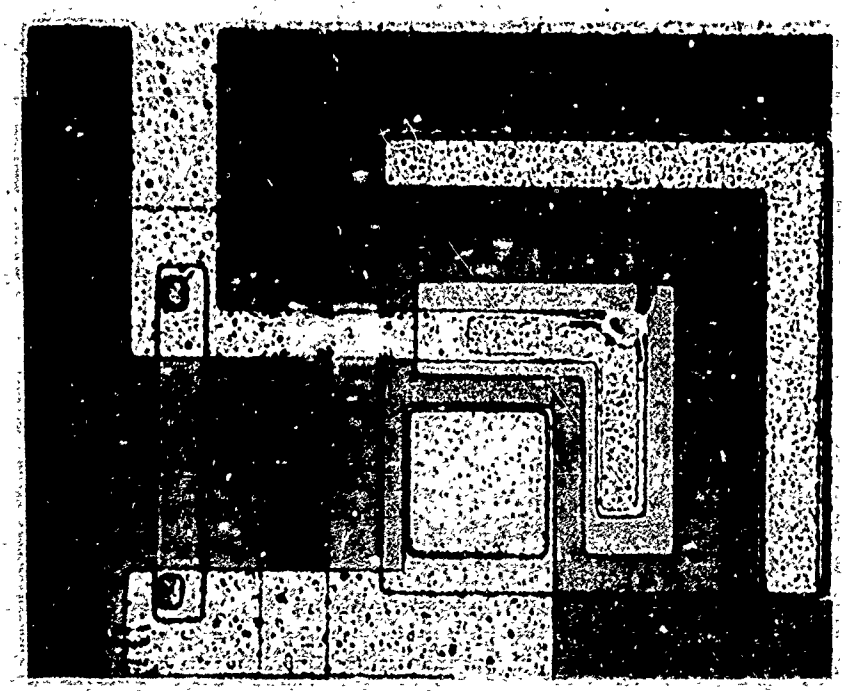


Fig. 51. Contact damage produced by pulses with second transition levels.

(c)



(d)

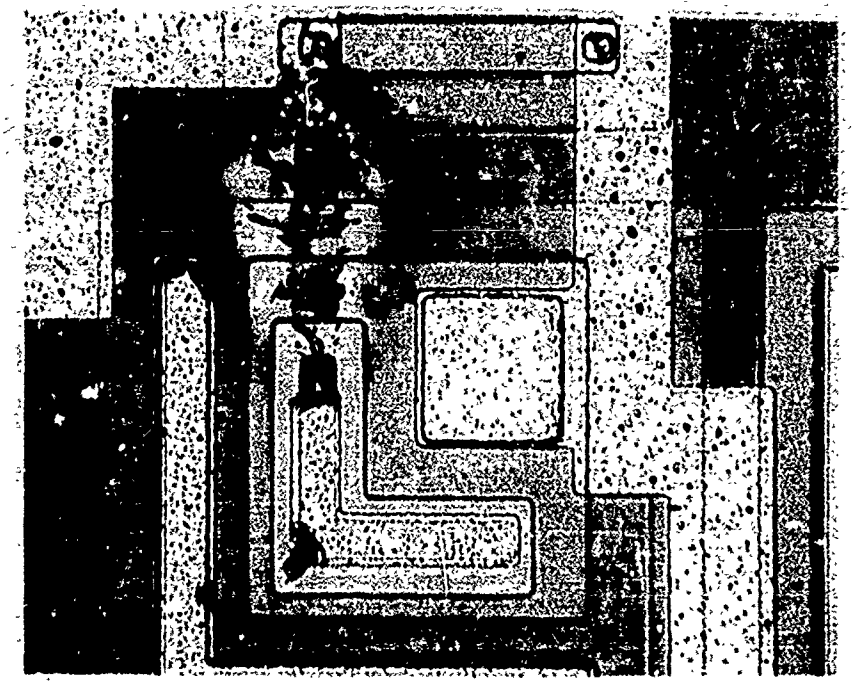


Fig. 51 (cont.). Additional examples of contact damage. Note the damage to metallization runs in each case.

Figure 51(c) shows damage from a 15 μsec pulse terminated just after the final voltage drop. A corner of the metal contact has melted and the track extends all the way to the collector. Of special interest is the melted metallization stripe. Melting occurs only where the oxide is thickest and probably was initiated at an oxide step. On this particular device the aluminum thickness is 8,000 Å.

The last transistor in the figure has the aluminum completely melted and opened. The damage was produced by the pulse of Fig. 50(c). Damage can be seen at the collector and emitter contacts and in the emitter region. Again, there is no sign of damage to the junctions in the picture.

The threshold power curve for collector-emitter tests is plotted in Fig. 52. The average power at each point was computed graphically from the voltage and current waveforms. There was no dependence on transistor geometry. In some events the device failed too early in the pulse to make an accurate estimate of the delay time. In such cases the power was computed from the preceding photograph and the point entered as "not failed" in the figure. The threshold power is greater for the collector-emitter mode than any other, exceeding collector-base pulsing by 50% at 1 msec and by 400% at 1 μsec.

In interpreting the results of collector-emitter pulsing the pertinent questions are: what is the significance of each of the two voltage drops, and how does each relate to device damage? The first transition, although thermal in nature, does not result in a microscopic melt; if it did there would be permanent damage. It is always reproducible. Contact damage does result after such a transition, so it indicates that a localized, high temperature site has formed. The mere fact that the voltage is discontinuous

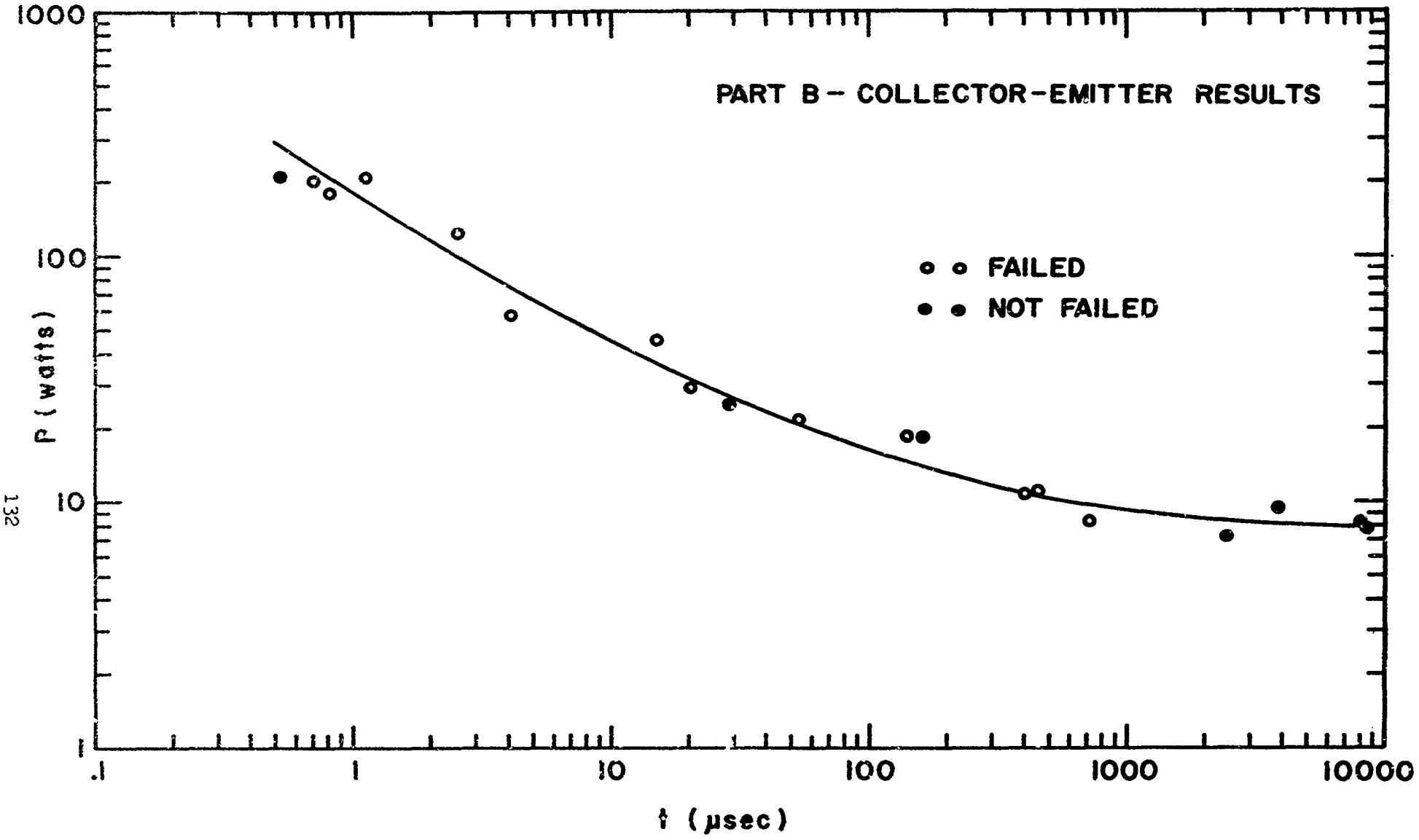


Fig. 52. Threshold damage curve for collector-emitter pulsing, Part B.

implies a change in the conduction mechanism.

The hypothesis is that the voltage drop signals the formation of a current constriction at the collector junction near the corner of the emitter contact and the formation of a broad filament extending from the constriction to the collector contact. The local temperature rises, but not to the eutectic point of the silicon-aluminum interface. This configuration appears to be stable for short periods of time if the total power dissipation is less than 15 watts. At levels less than this the voltage remains constant or rises almost imperceptibly for periods ranging up to 200 μsec. When the power level is greater than 15 watts the voltage continues to fall slowly. After a delay period the process reaches the second phase.

The final transition introduces permanent damage. This means localized melting has occurred. The proposition is that the filament grows from the collector junction through the base region, across the forward biased emitter junction and to the emitter contact. Then a melt forms.

The extent of the damage would depend primarily on how long the pulse lasted after breakdown occurred and on the power levels. In every case the damage at the emitter contact was apparent. Occasionally, localized damage appeared at both collector and emitter without a stress mark in between (as in Fig. 51(d)). This is an indication that a filament extended from contact to contact. If the oxide and aluminum had been removed on these devices it seems highly likely that a continuous damage streak would be found.

This model can only be regarded as speculative since no etching or cross-sectioning of these devices was done to determine the location and extent of the resolidified melt. Nor is there any way to determine the sequence of steps as in the SOS diode experiments. It does fit the observed

facts for this transistor, however, and is a reasonable extension of second breakdown in simple junctions. Budenstein, Pontius, and Smith have shown that broad filaments can exist and remain stable without melting under certain situations. In the results of the next device to be presented there is evidence that transitions similar to the final voltage drop are associated with damage in the emitter region. Thus, although the model cannot be regarded as conclusive, there is good evidence to support it.

Another theory that has been proposed is that the final, destructive phase is due to the melting of aluminum at the contact which then alloys into the silicon to form a melting, conducting channel. This does not appear to be the case here. There is evidence that the voltage transition must be essentially complete before permanent damage is encountered. If melting aluminum initiated the final phase, some form of damage would appear even if the transition was not complete.

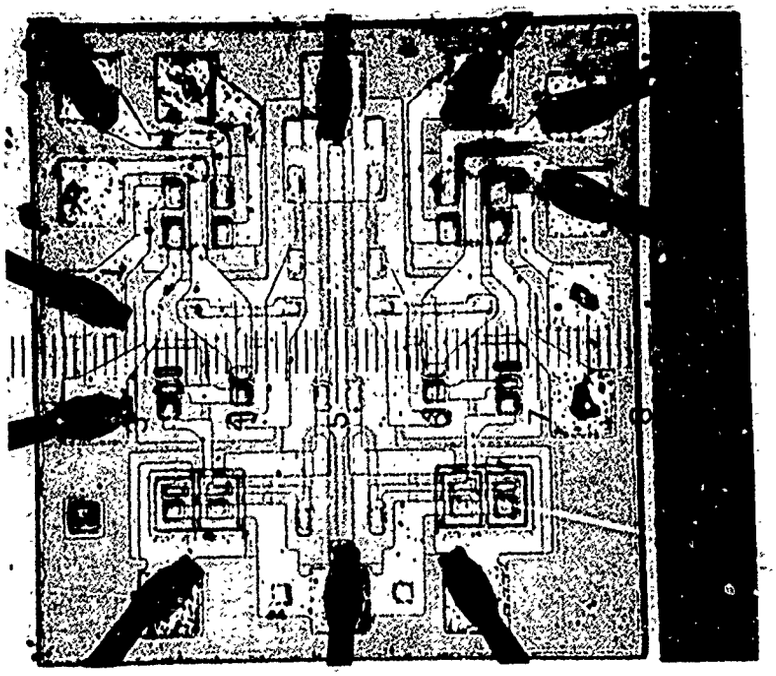
3. Part C

This circuit is a standard DTL NAND gate. Photomicrographs of the circuit appear in Fig. 53. The chip is 32 mils square and is glassivated. The input diodes, in the upper right and left corners of the picture are the only components which can be individually tested. Part (b) shows details of one input group. Each division is 2.5 μm.

The 4 diodes are formed by emitter diffusions into two separate base regions. The two p-regions are connected by an overlapping n^+ diffusion, and an aluminum anode stripe makes contact to the n^+ region and the p-type regions on either side. A diagram of this construction and the pin connections are given in Fig. 54.

Three different diode configurations were tested, all at room

(a)



(b)

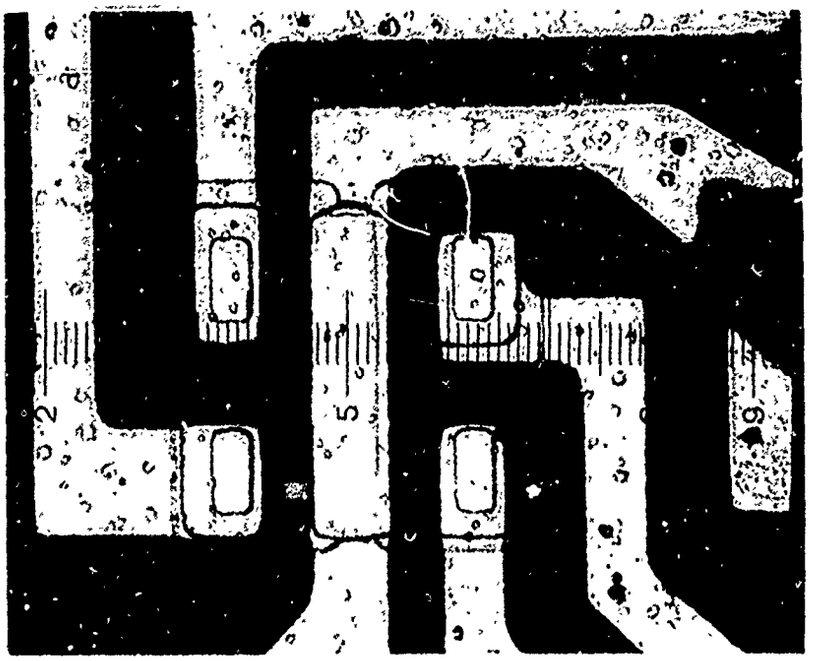


Fig. 53. Photomicrographs of the circuit of Part C. (a) Each division is 12.5 μm. (b) Close-up of the diode cluster. Each division is 2.5 μm.

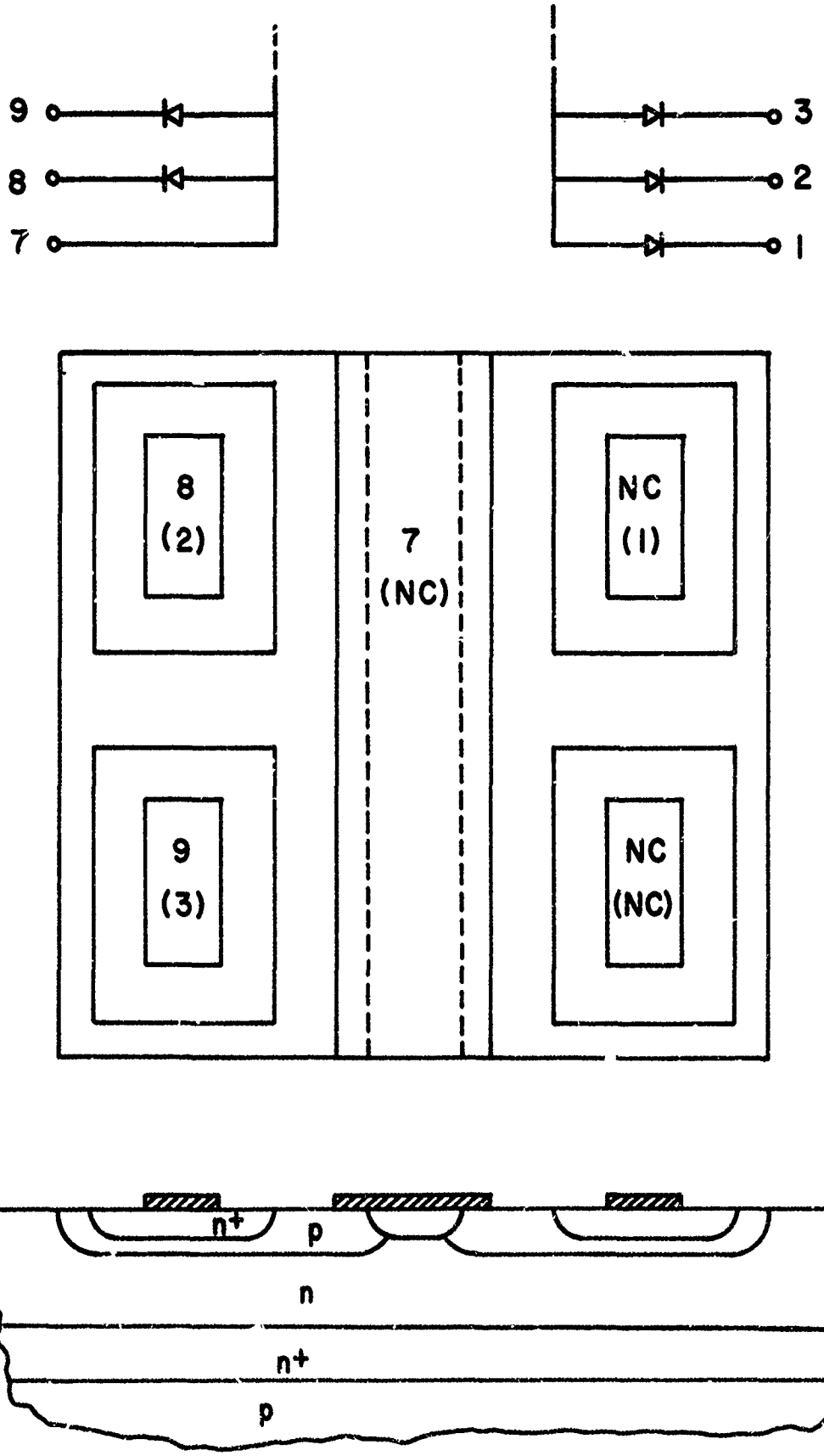


Fig. 54. Pin connections for the input diodes and details of the construction.

temperature and all in either the reverse bias or the back-to-back mode. The standard configuration is simply an ordinary diode connection between pin 8 or 9 and pin 7. The second combination is two diodes on opposite sides of the anode stripe connected back-to-back (pins 1 and 8). For the third combination two diodes in the same p-region (pins 2 and 3) were pulsed back-to-back.

The diode characteristics are shown in Fig. 55. Part (a) is the characteristic between pins 9 and 7. In the forward direction the diode shows signs of reaching high level injection when the current is only 15 ma. In the reverse direction a discontinuity is seen as a result of the interaction between the emitter and collector junctions of the transistor formed by the n^+ diffusion, the p diffusion, and the n-type isolated region. In other words the reverse characteristic shows the collector-emitter breakdown of the transistor operating in the inverted mode.

Part (b) is a back-to-back diode characteristic. Part (c) shows a portion of the reverse breakdown characteristic on an expanded scale to show the difference in voltage drops across different pairs of diodes. The characteristic with the smallest voltage drop is for two diodes in the same p-region, while the other trace belongs to two diodes on opposite sides of the center contact. The increased voltage drop had little effect on threshold power except at short pulse lengths.

Representative second breakdown waveforms are shown in Fig. 56. The pulse in the first photograph was applied between pins 1 and 2. Waveforms for this connection were typically of this form - a relatively small, slow voltage drop often in two steps, with a large residual voltage. Visible damage was not often apparent in this mode of pulsing. When damage could be

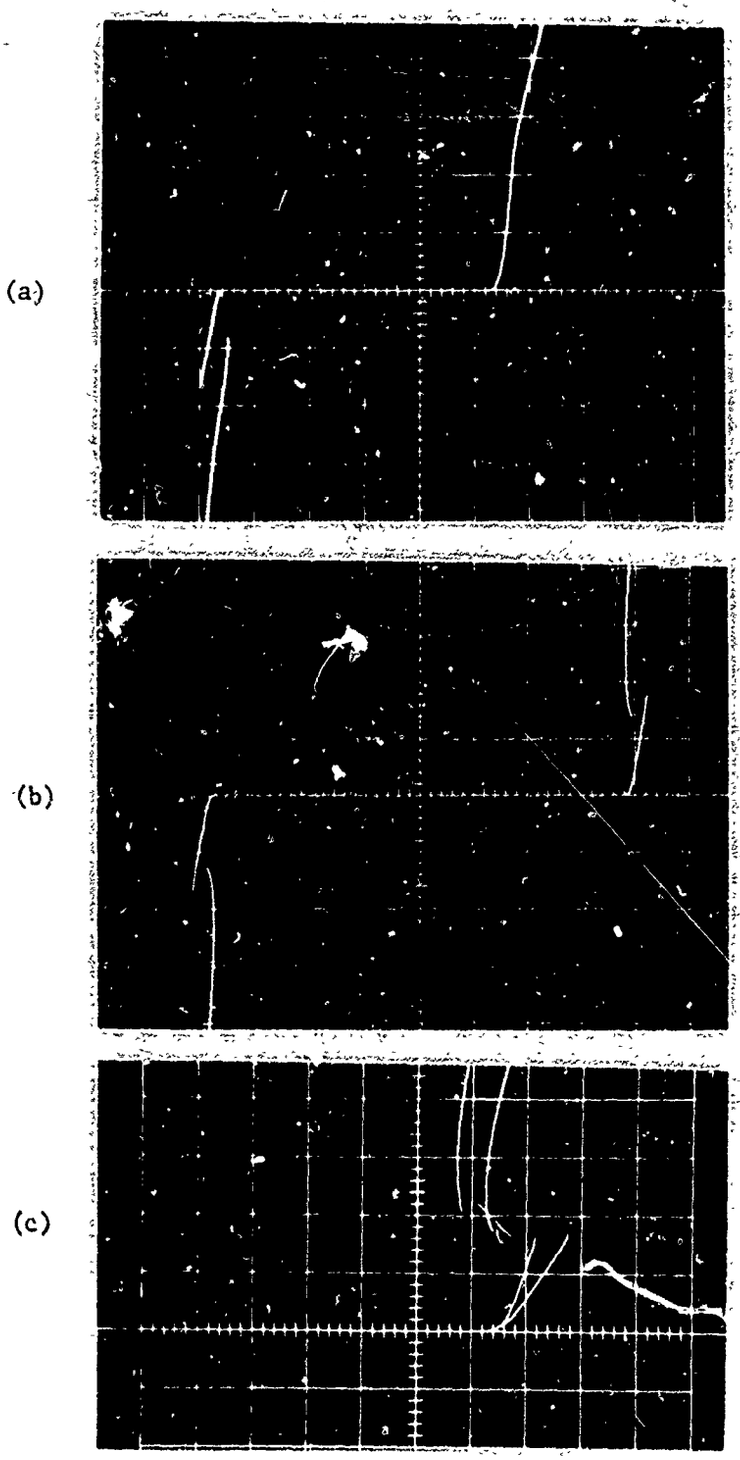


Fig. 55. Diode characteristics. (a) Pins 9 and 7. Forward; 0.5 v/cm, 5 ma/cm. Reverse; 2 v/cm, 5 ma/cm. (b) Pins 9 and 8. 2 v/cm, 5 ma/cm. (c) 0.5 v/cm, 5 ma/cm. Left-hand characteristic, pins 2-3. Right-hand characteristic, pins 1-3.

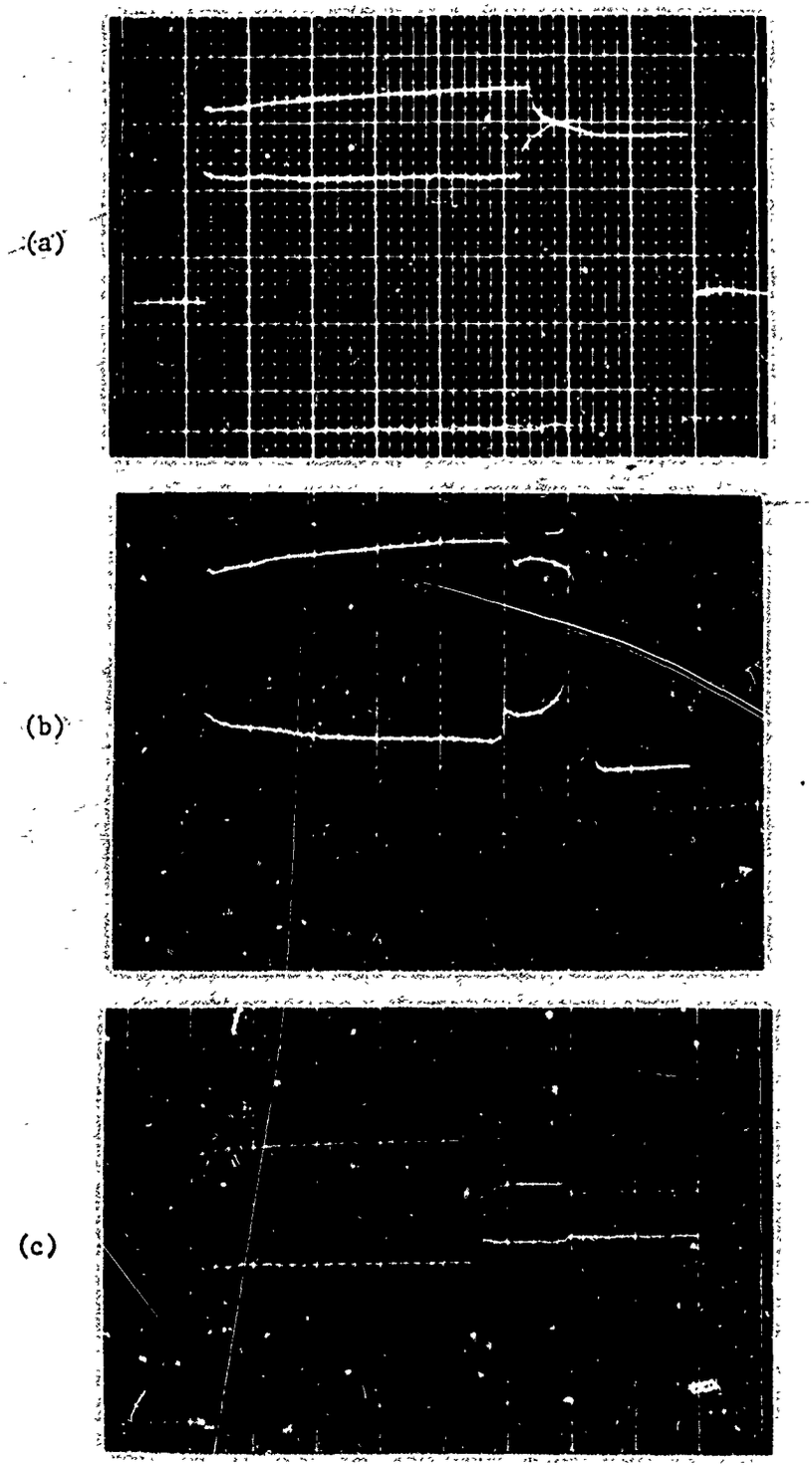


Fig. 56. Breakdown waveforms. (a) Pins 1-2. 5 v/cm, 100 ma/cm, 2 μsec/cm. (b) Pins 8-7. 5 v/cm, 100 ma/cm, 2 μsec/cm. (c) Pins 3-2. 5 v/cm, 100 ma/cm, 10 μsec/cm.

seen it appeared at the contact and was localized. The glass layer and aluminum were removed on one part. When a Dash etch was applied the zap mark could be easily discerned extending downward and into the emitter material. The exact extent of the melt column could not be determined. However, there was never any indication of damage extending laterally into the p-region except on one part which failed from repetitive pulsing.

Microsecond-length pulses applied to any of the three diode configurations resulted in breakdown waveforms very much like the one in part (a) of the figure. Any visible damage was always confined to the n^+ emitter region in such cases.

Longer pulses applied to pins 1 and 2 produced the same type of waveform and the same type of damage. Longer pulses applied to the other two diode configurations produced different results. The waveform in part (b), Fig. 56 is typically seen between pins 8 and 7. The first transition does not produce permanent damage. The second transition does. Furthermore, whenever both transitions appear the zap mark extends past the emitter junction, through the p-region, and to the aluminum contact. In other words whenever a pulse produces permanent damage a zap mark can always be found in the emitter material. Whenever damage is found in the base region, both transitions are present in the pulse waveform. It appears then that the first transition in part (b) of the figure is caused by filament formation in the p-region. The second voltage drop comes about when the filament extends through the emitter region to the emitter contact and a melt forms. Thus the breakdown and damage process is similar to that of the transistors of Part B.

The third photograph in Figure 56 is the waveform between pins 2

and 3. The emitter regions are lined up end to end in this configuration and the power levels are lowest. Again the double transition is associated with damage to both the n^+ and the p-region while a single transition is associated with damage in only the emitter region. It appears possible that the order of events could be reversed in part (c), but there is no way of determining this from the data which was taken.

Figure 57 shows the threshold damage curves for this circuit. Due to the small size the power levels are low, and the t^0 asymptote is reached in about 100 μsec. The lowest curve is for pulses applied between pins 2 and 3. In the other configurations the emitter diffusions are broadside to each other and the power levels are 50% greater. This points out that the power required to damage a circuit depends on the length of the active portion of the emitter junction and not on the area of the emitter diffusion.

The figure also shows that the threshold power is the same for connections 1-2 and 8-7 for t greater than 10 μsec. One would expect at least a small difference. No matter whether the current path between 1 and 2 is along the surface or via a buried subcollector there are extra parasitic losses as the characteristics of Fig. 55(c) show. The voltage change is only about 4% at low current levels and if it is about the same for low power pulses, it would be difficult to see in the normal 10-15% scatter of the data points. At high pulse power levels the increase in parasitic losses is readily apparent. The damage curve rapidly approaches a t^{-1} variation, indicating power dissipation throughout a broad region.

In most cases damage to the devices - both electrical and visible - was slight. Electrical damage often consisted only of small changes in the knee of the breakdown curve. Long pulses caused greater damage but not so

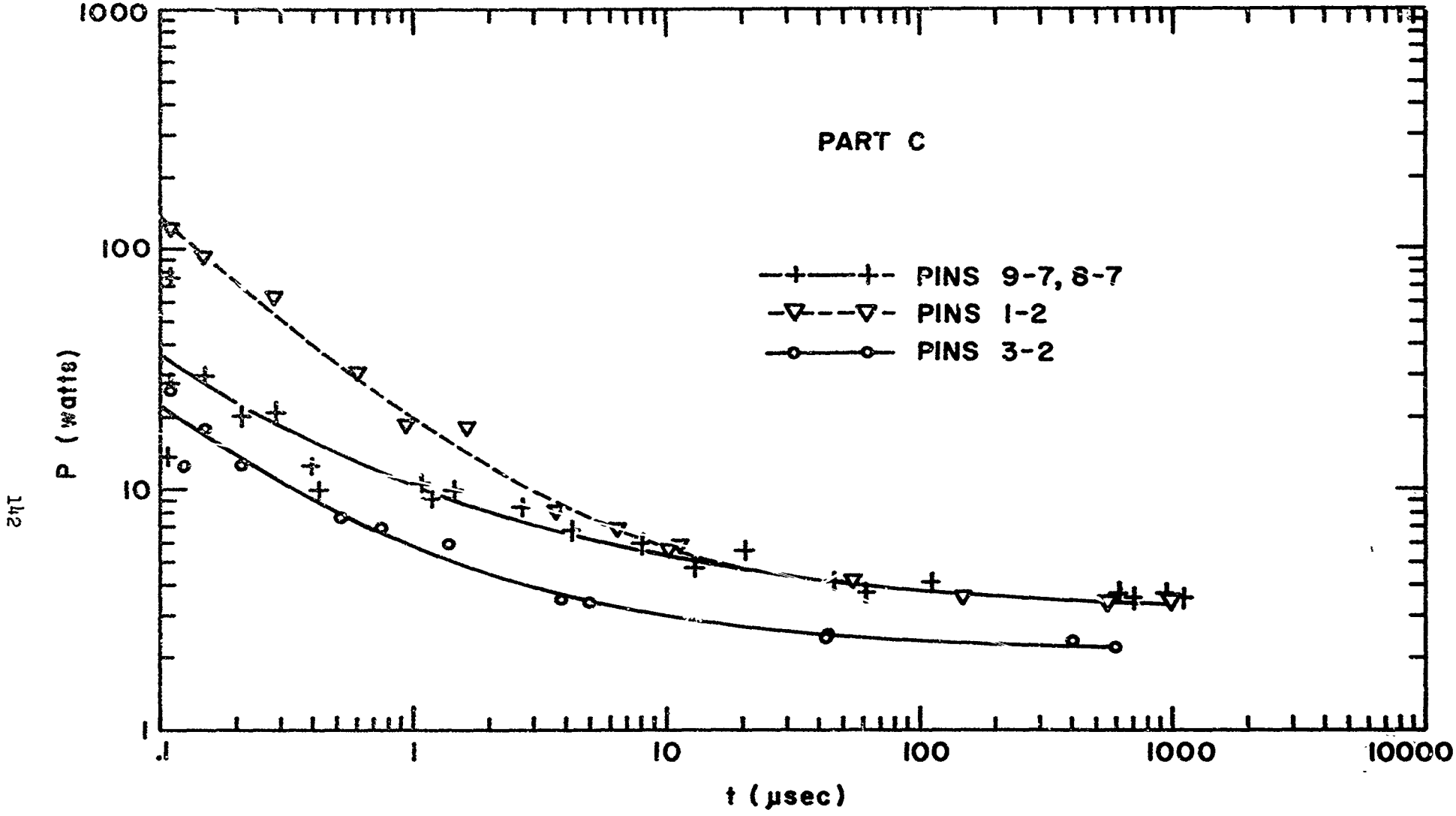


Fig. 57. Threshold damage curves for the three diode configurations of Device C.

much as to give the appearance of a low resistance short across the junction. Zap marks were similar in appearance to those of other base-emitter junctions except that damage always existed in the emitter region and sometimes only in the emitter region. There was rarely any damage to the metallization. One reason was the small amount of heat-sinking by the glass layer (about 15 μm thick). A more important reason was the small size of the devices and the lower total power levels involved.

The upshot of all this is to point out the differences in second breakdown between simple p-n junctions, integrated circuit transistors, and three-region structures such as those discussed here. In all cases, a constriction forms at the reverse biased junction. With the exception of Part C under high power pulses, constriction is followed by filament formation in the high resistivity material. In a simple p-n junction the next stage is melt formation and permanent damage. In other devices the filament grows all the way to the emitter contact before damage occurs. At high power levels in the Part C structures the filament forms only in the emitter region before failure occurs.

This model cannot be regarded as conclusive, but there is good correlation among the various devices tested in spite of the differences in size and construction.

4. Part D

Part D is a dielectrically isolated dual DTL expandable NAND gate. 4 diodes with a common anode terminal are fabricated on each side of the chip. A photomicrograph of two of the input diodes is shown in Fig. 58.

The diodes are fabricated in pairs in four separate isolated n-regions. All diodes have the same dimensions but each pair has a different

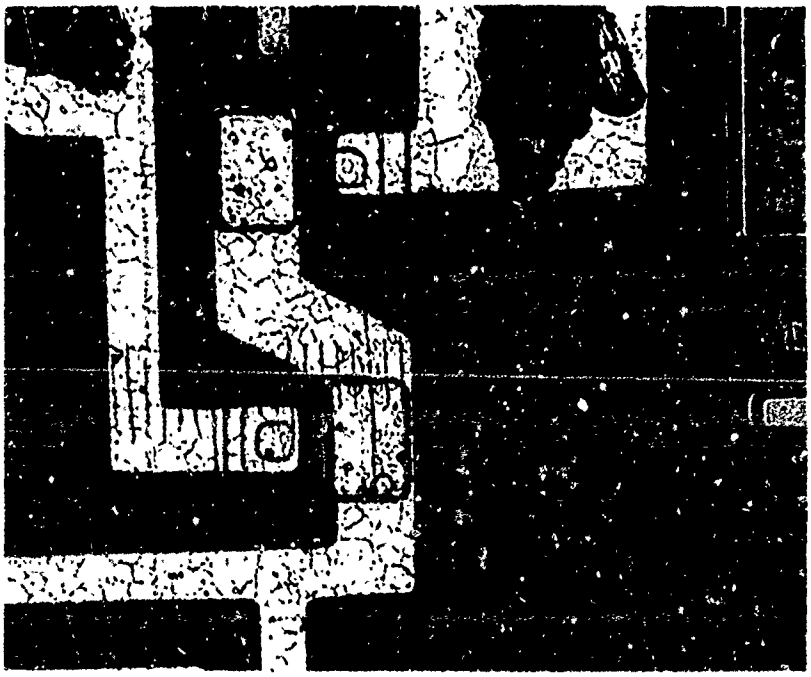


Fig. 58. Part D input diodes. The chip utilizes dielectric isolation. Each division is 6.25 μm.

arrangement within the n-region. Each diode is formed from a p-type diffusion with an n^+ diffusion serving as the cathode. A separate n^+ diffusion overlaps part of the p diffusion and the n-region. Aluminum deposited on top of this serves as the anode contact. The base sheet resistance is 110 ohms per square and the breakdown voltage is 7.2 volts. Once again it is really the breakdown of an n^+-p-n structure operated in he inverse mode. A discontinuity in the breakdown characteristic occurs at 10 ma.

Second breakdown waveforms are shown in Fig. 59. They are distinctive because of the relatively small voltage drop during second breakdown. For pulse widths less than 1 μsec the diodes were checked on a curve tracer after each pulse to determine whether or not damage had occurred. For pulse widths greater than 100 μsec the voltage waveform usually showed a second small discontinuity such as appears in part (c) of the figure.

In a few instances the second breakdown transition was not accompanied by electrical or visual damage. Whenever this occurred a succeeding pulse of higher amplitude caused breakdown at a level comparable to that of an undamaged junction.

Visible damage was either not apparent or appeared as a small mark in the emitter region. Strangely enough the mark sometimes appeared at the edge of the diode rather than in the region between the contacts. No damage was ever seen to the metallization during single pulse testing. Two diodes were stressed in the repetitive pulse mode while being monitored with the closed circuit TV system. These were the only diodes where well defined stress marks could be seen extending from contact to contact and where damage was seen to the contacts themselves.

The second breakdown power versus delay time curve is plotted in

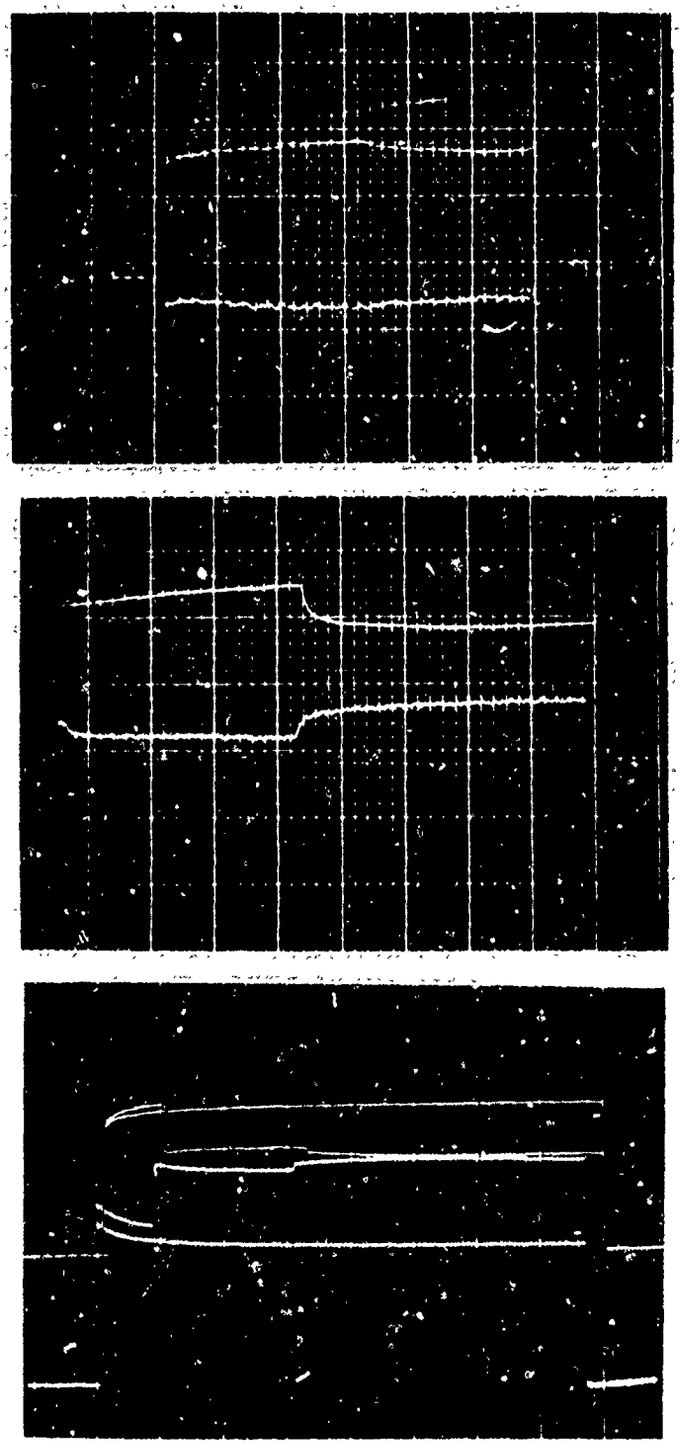


Fig. 59. Breakdown waveforms. (a) 10 v/cm, 500 ma/cm, 200 nsec/cm. (b) 5 v/cm, 200 ma/cm, 1 μsec/cm. (c) 5 v/cm, 100 ma/cm, 100 μsec/cm.

Figure 60. Note the well defined t^{-1} dependence for t less than 1 μsec. There is little scatter over the whole range. In spite of the fact that these diodes are the smallest tested thus far the power levels are not the lowest.

5. Part E

This circuit is a dual 4 input diode expander fabricated on a dielectrically isolated and glassivated chip. The diodes are emitter-base diodes, arranged in groups of 4, two each in a separate p-type region. An aluminum metallization run makes contact to two adjacent p-regions and also connects the n-type region. A layout of a diode group is given in Fig. 61. The p-diffusion is about 2.6 μm deep and the n-type tub is about 13 μm deep. The diodes have a reverse breakdown voltage of 7.2 volts and show the familiar discontinuity due to n^+-p-n action.

About 30 diodes were tested, all in the reverse direction and all at room temperat The only configuration which was tested was between the emitter diffusions and the common aluminum contact.

Several typical breakdown waveforms are given in Fig. 62. These are quite similar to the waveforms of previous circuits. The voltage drop is comparatively small in every case. The third photograph shows three successively higher pulses. The third pulse causes breakdown and shows the small post-breakdown drop in voltage commonly seen in previous devices.

In all but three cases the second breakdown transition produced permanent changes in the diode characteristics. A systematic investigation to determine the nature of the damage was not undertaken, but on two packages which were opened the damage was much like that of diodes tested earlier. Such damage as could be seen was minor, consisting of only a small mark in the emitter region and no contact damage.

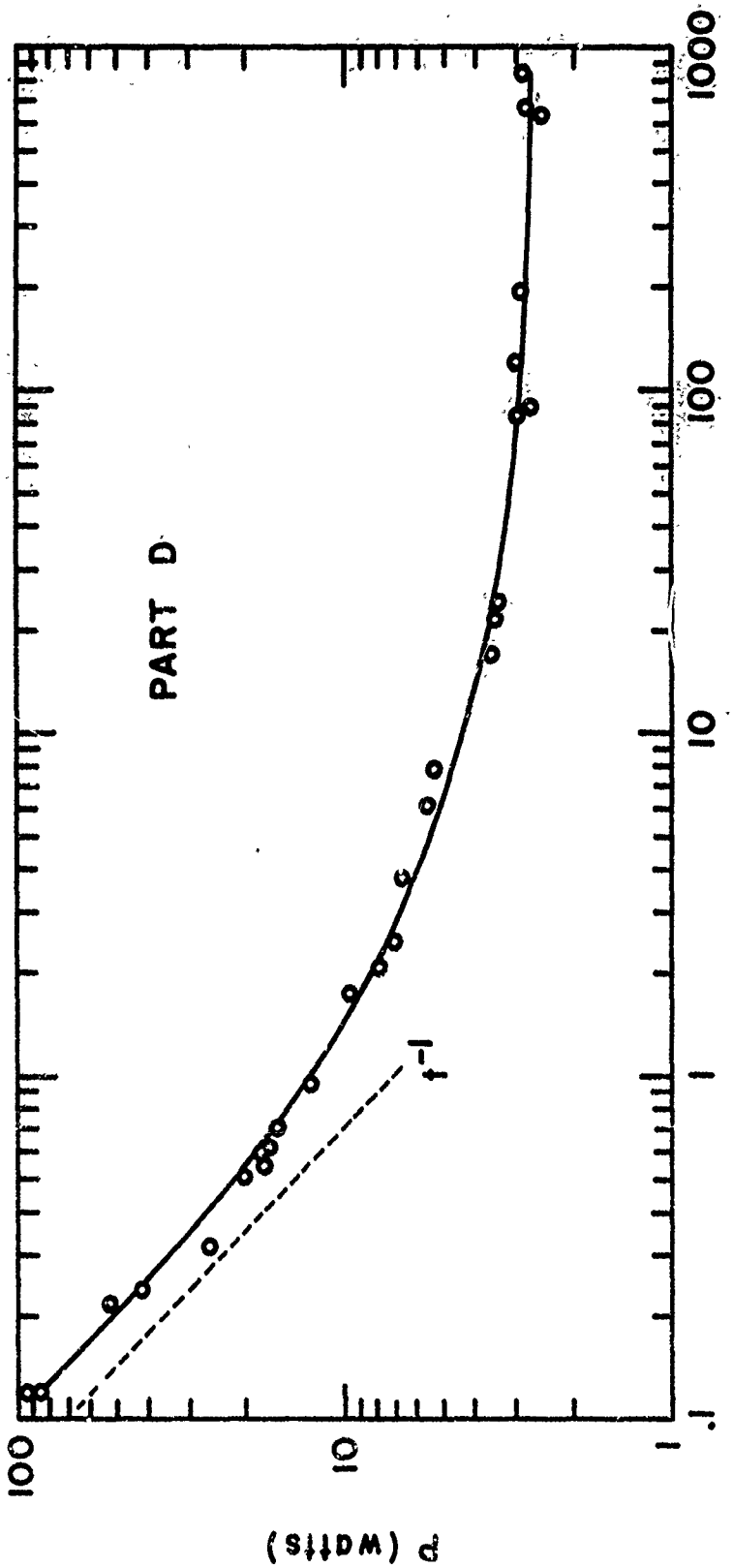


Fig. 60. Threshold damage curve for Part D.

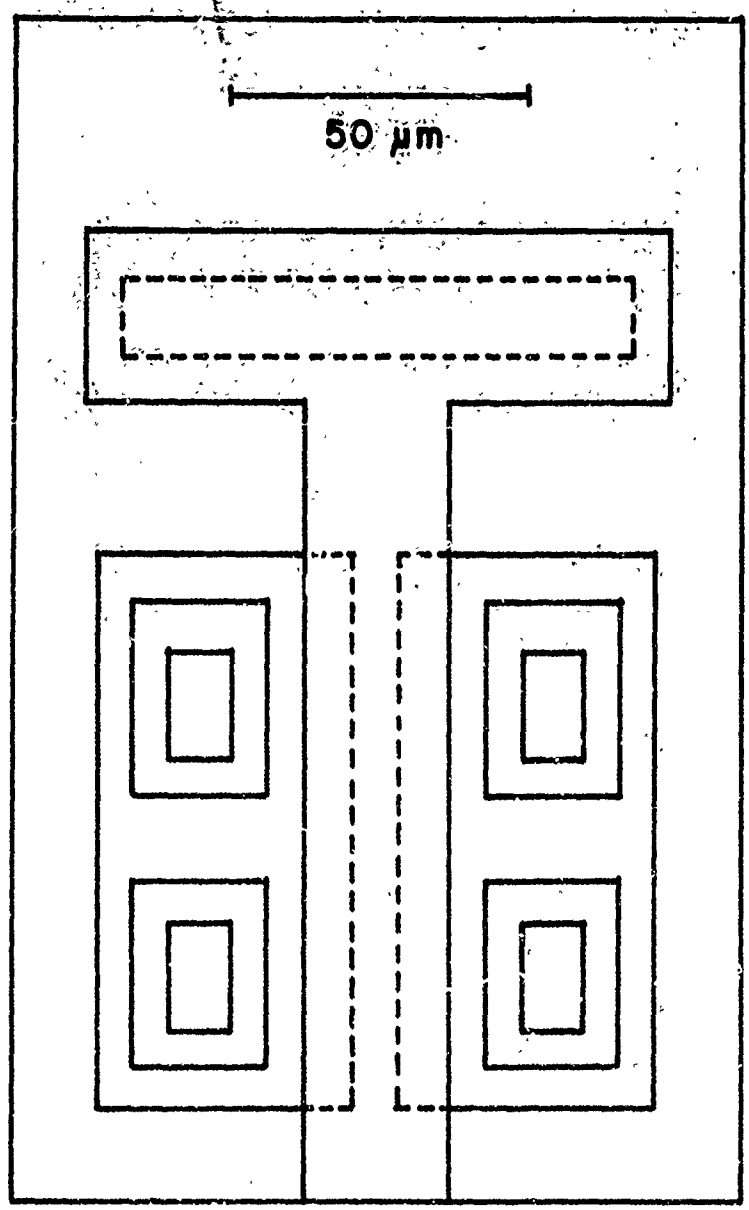


Fig. 61. One of the input diode groups on Part E. The T-shaped metallization is the expander terminal and serves as the common adode for the four diodes, making contact to both p-regions and the n-type isolation region.

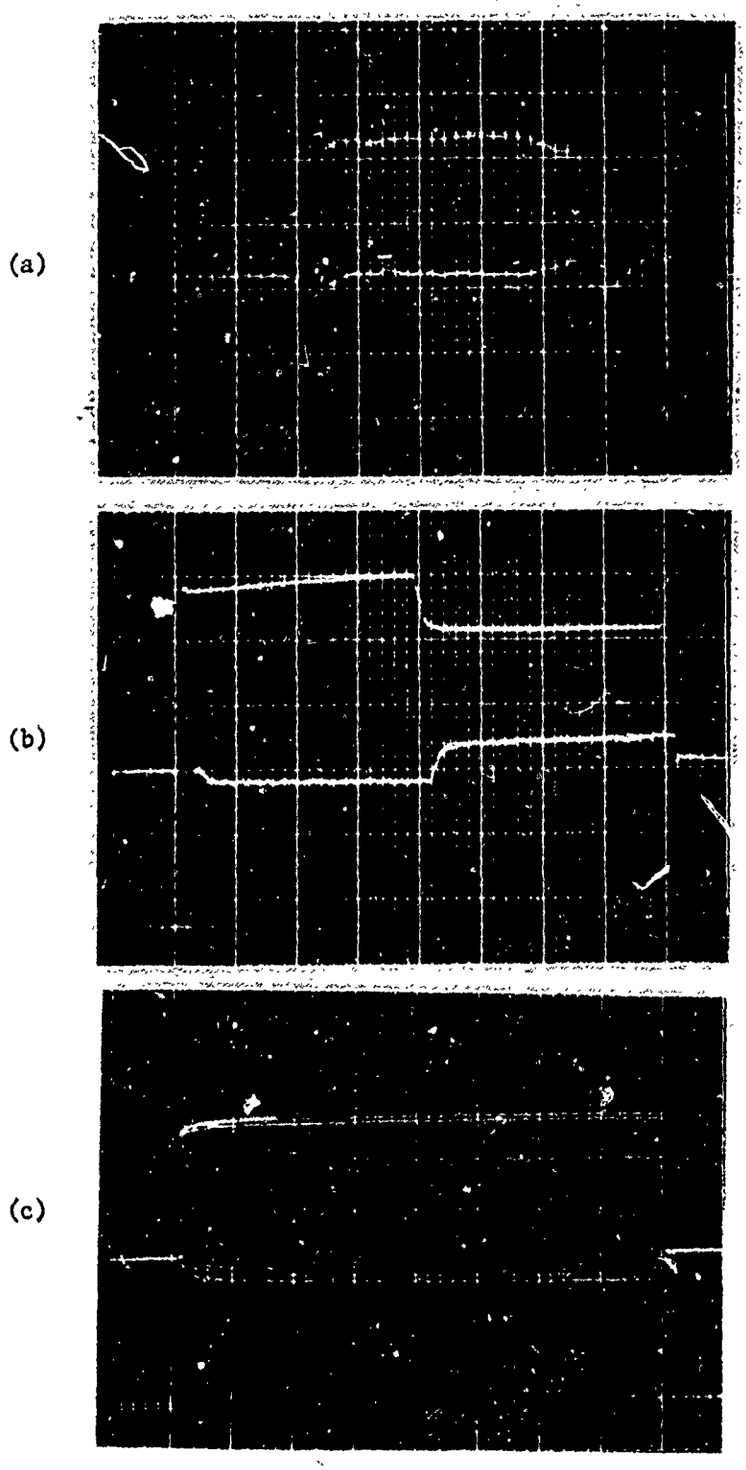


Fig. 62. Breakdown waveforms for the diodes of Part E.
(a) 10 v/cm, 500 ma/cm, 100 nsec/cm. (b) 5 v/cm, 200 ma/cm, 1 μsec/cm. (c) 5 v/cm, 100 ma/cm, 50 μsec/cm.

The threshold damage versus pulse width curve appears in Fig. 63. The power approaches a constant value for t greater than 100 μsec as expected for a small device. At the other extreme, for t of the order of 100 nsec, the expected asymptote is not yet reached. This is probably due to normal variations in threshold power and is not necessarily significant in itself.

6. Part F

The final circuit tested was a dielectrically isolated quad core driver. There were four large interdigitated transistors on each chip. All terminals were individually connected, making it possible to perform tests on all three configurations. Each transistor measured roughtly 20 mils by 40 mils. The manufacturer's specifications are as follows, with the measured values in parentheses:

BV_{CEO}	40 volts	(43)
BV_{CBO}	80 volts	(120)
BV_{EBO}	6 volts	(7)
h_{FE}	50	(90)
P_D	800 mW at 25°C	

The data obtained during this portion of the experimental phase is somewhat meager for several reasons. Only 7 parts were available. The large size of the transistors indicated at the outset that sufficient pulse power was probably not available at short pulse widths. This was found to be true although some devices failed at pulse widths less than 1 μsec. Even more important, the threshold power was erratic from one transistor to another. When a failure point was determined for one transistor, others pulsed under the same conditions would fail prematurely resulting in little useable data. Finally, when a failure was induced, it was nearly impossible to locate the

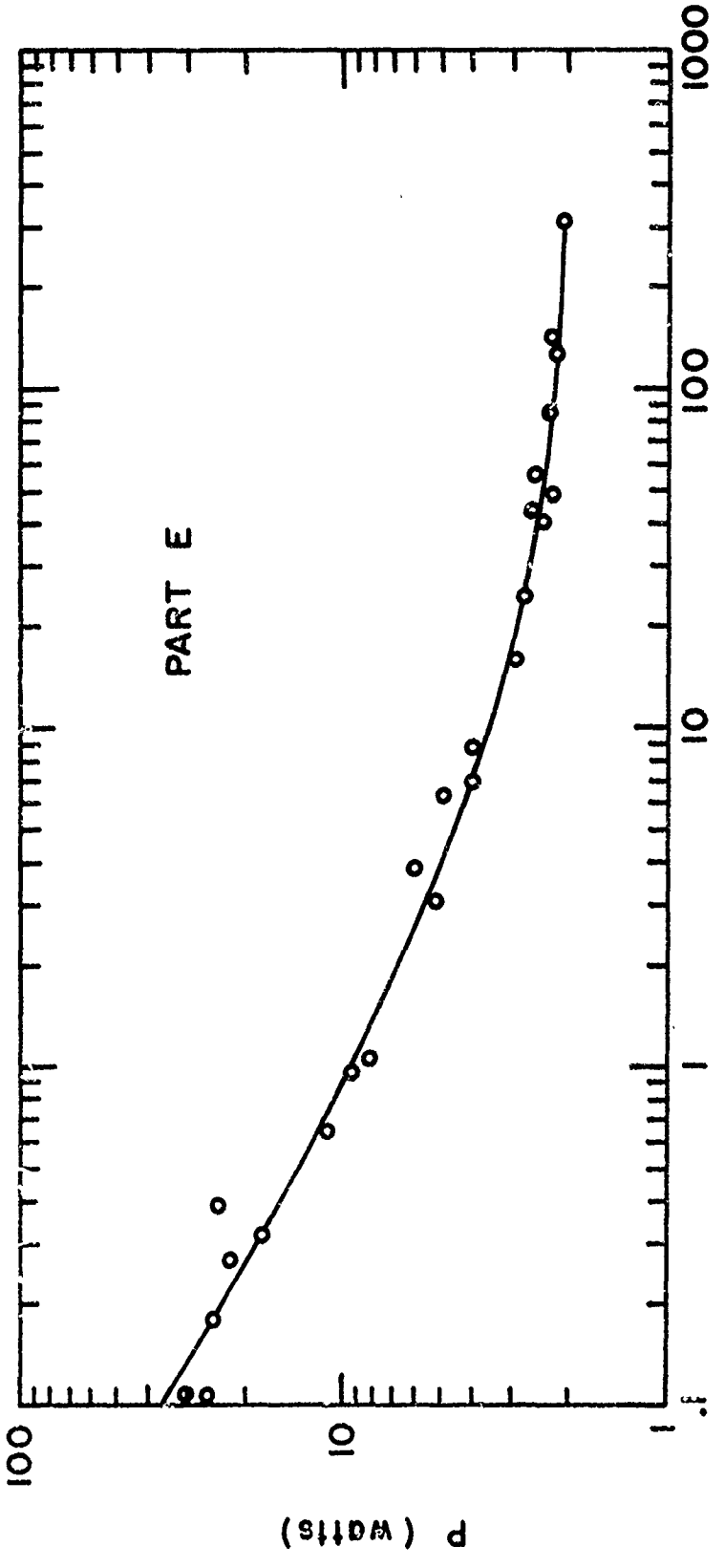


Fig. 63. Threshold power curve for Device E.

site of damage in the vast expanse of emitter and base fingers.

Contrary to earlier testing procedure the method used here was to select a pulse amplitude and then increase the pulse width until damage occurred. Sixteen transistors were pulsed to failure in this manner in the collector-base mode. Eight others were pulsed collector-emitter and four emitter-base.

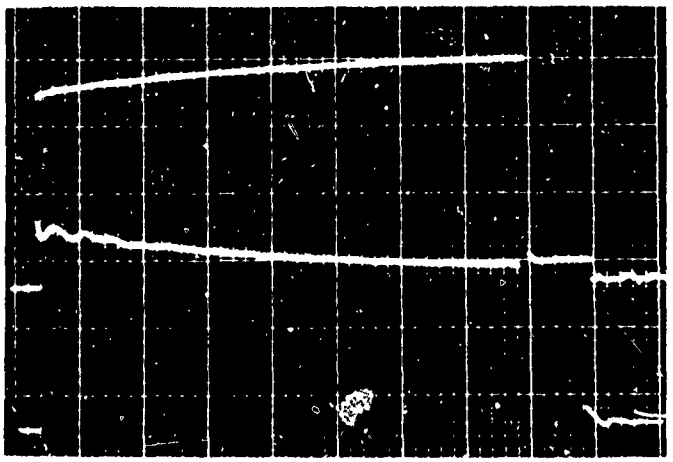
Breakdown waveforms appear in Fig. 64. The first picture shows collector-base breakdown consistently observed for pulses longer than 10 μsec. The voltage rises slowly and the current drops correspondingly with a "time constant" of roughly 40 μsec. Second breakdown manifests itself in a sharp drop to a low voltage. Damage always followed this transition.

Three transistors on one chip failed in less than 200 nsec. All three waveforms showed a voltage drop from 130 volts to 50 volts, then a further drop to 20 volts.

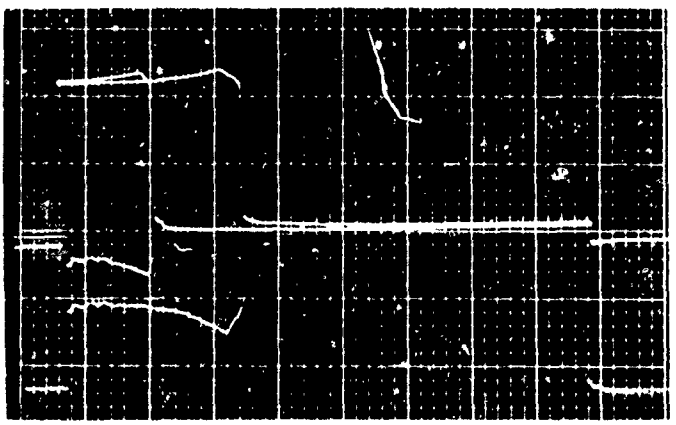
Collector-emitter pulsing yielded a variety of waveforms. At short pulse widths second breakdown was not apparent from the waveforms and the transistors had to be checked after each pulse. At intermediate pulse widths breakdown sometimes occurred in two separate, distinct steps of equal magnitude. This is reminescent of collector-emitter breakdown in the transistors of Part B.

At long pulse widths the waveforms had a different appearance. An example appears in part (b) of Fig. 64. Two different pulses are shown. The voltage in each case rises to a peak, decreases slowly, then drops abruptly. More important, both pulses were applied to the same transistor. This means that damage was not detected on the curve tracer after the first pulse. In the collector-emitter connection it was often possible to pulse

(a)



(b)



(c)

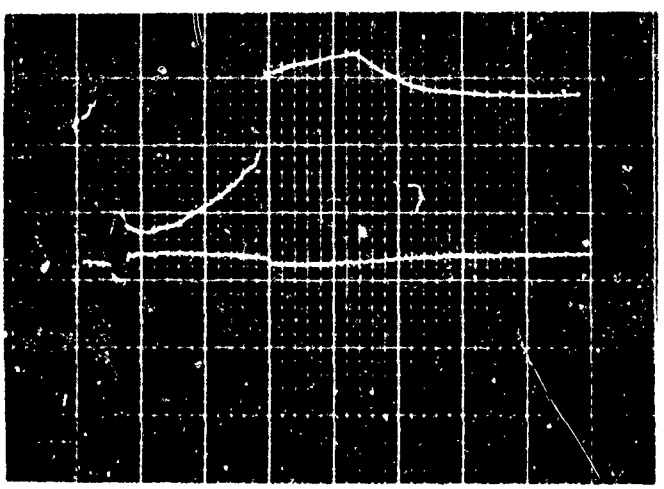


Fig. 64. Breakdown waveforms for core driver transistors.
(a) Collector-base breakdown, 50 v/cm, 200 ma/cm, 10 usec/cm.
(b) Collector-emitter. 20 v/cm, 200 ma/cm, 20 μsec/cm.
(c) Emitter-base. 5 v/cm, 2 A/cm, 1 μsec/cm.

the transistor more than once into second breakdown before degradation was apparent. Conceivably, a small damage site might go undetected because of the large size and current carrying capability. This is not likely though because the threshold on subsequent pulses was in accord with an undamaged junction.

Part (c) of Fig. 64 is the waveform of emitter-base breakdown. All four transistors on one die showed the same type of waveform for a total pulse width in the range from 4 to 80 μsec. Damage was observed in each case following the appearance of this unusual waveform.

Threshold power points are plotted in Fig. 65. The data points are so dispersed as to make it impossible to draw meaningful curves. Power levels are lowest for the collector-emitter points, contrary to the transistor results of device B. The only conclusion which can be drawn from the figure is that large area components have a wide variation in breakdown thresholds.

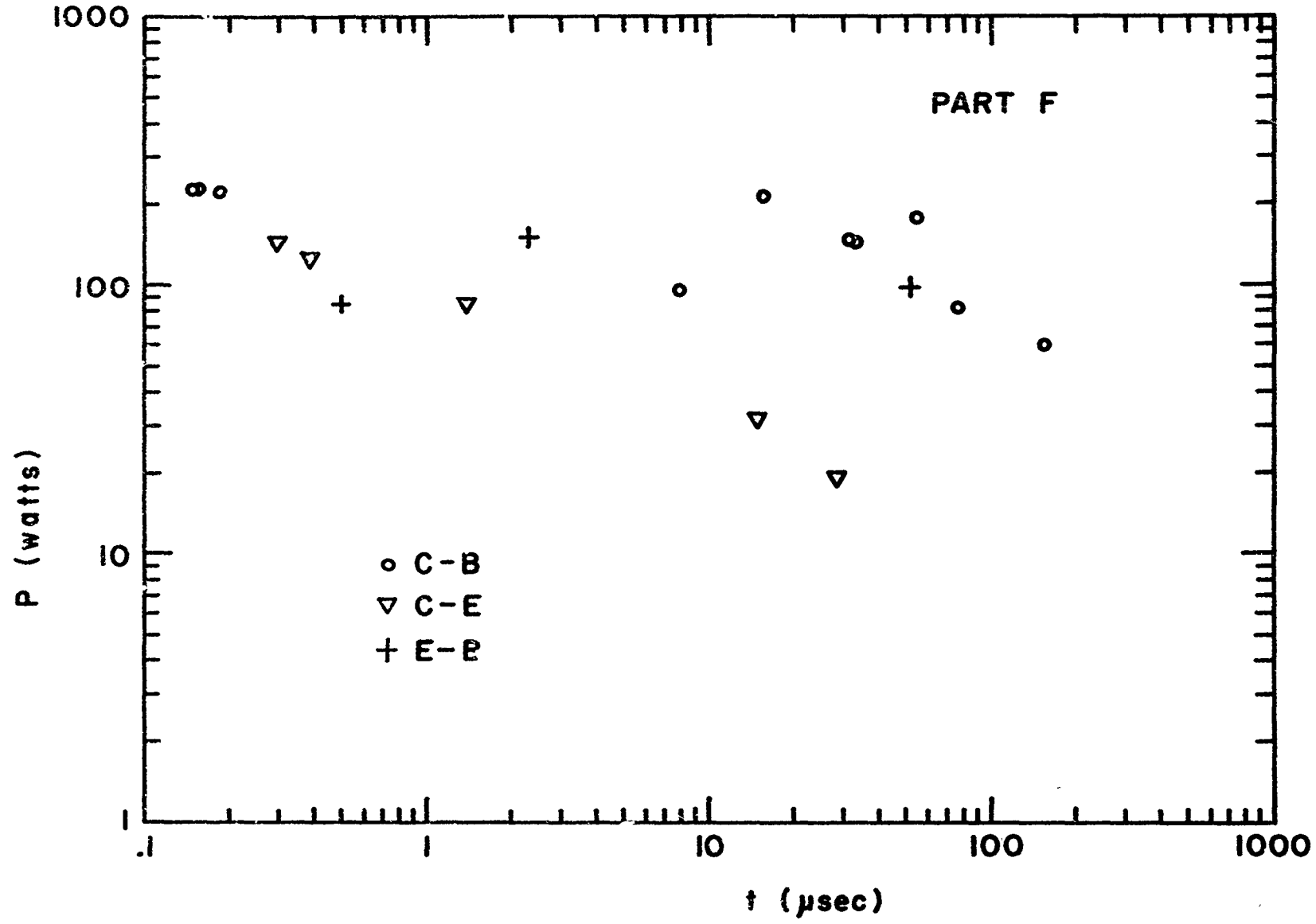


Fig. 65. Threshold damage points for the core driver transistors of Part F.

SECTION V

DISCUSSION

Since current constriction, filamentation, and second breakdown occur at temperatures where either the material becomes intrinsic or the thermally generated currents quench the avalanche breakdown, the temperature rise of the silicon is of major importance in determining the outcome of a particular experiment. If the power distribution is known then the temperature rise throughout the chip can be calculated by techniques such as those described in Section II. Unfortunately, a change in temperature causes a redistribution of the dissipated power and an exact analysis becomes very complex.

In a reverse biased diode the greatest power density occurs at the depletion layer. Some of the pertinent questions relating to this portion of the total power are: how wide is the depletion layer; how is the voltage related to doping levels and junction curvature; how does the voltage change with temperature; and how is the voltage affected by high current densities?

For emitter junctions most of these questions can be readily answered. The high doping levels result in a low breakdown voltage, typically between 6 and 7 volts. Calculations not included in this report indicate that the depletion layer in such cases is very narrow -- of the order of 0.1 μm. This thickness is much smaller than the radius of curvature of a typical emitter diffusion; hence, the breakdown voltage is not limited by junction curvature. Avalanche breakdown usually occurs at the surface because of the heavier base doping at the surface. This localization of breakdown is further enhanced by lateral IR drops in the base region which

produce an effect like current crowding. The upshot of all this is that the current flows across the junction at the edge of the emitter closest to the base contact.

Temperature variations of the breakdown voltage should be nearly zero in emitter-base junctions, unlike high voltage junctions where the change in ionization coefficients produces a positive temperature coefficient. Consequently, temperature changes would be expected to have little effect in themselves on the power dissipated at the junction. This was not found to be entirely true as will be explained below.

Changes in junction voltage and depletion layer width can occur at high current densities when the carrier densities in the space charge layer become comparable to the densities of donors and acceptors. The space charge resistance which arises under these circumstances can be either positive or negative, depending on the impurity distribution. For an abrupt junction the space charge resistance is positive, but for a p-i-n diode it is negative.

The remainder of the applied power is dissipated in the diode bulk resistance, contact resistance, lead wire resistance, etc. If the geometry and doping levels are known, the parasitic resistance can be readily estimated.

When the above considerations are directed toward an analysis of the diodes of Device A it becomes apparent that the greatest power density does indeed occur at the junction but that the largest portion of the total power is dissipated in the parasitic base resistance. The series resistance calculated from the sheet resistivity is about 25 ohms. Voltage and current were read from the leading edge of pulse waveforms for voltages ranging up to 30 volts and current up to 1 amp. For diodes with a 10 μm base region

these points fall on a straight line with an incremental resistance between 25 and 30 ohms. Furthermore, the resistance depends on the junction to base contact distance. For a diode with a base width of 14 μm the incremental resistance is between 35 and 40 ohms, decreasing to 15-20 ohms for a 5 μm base width.

At low current densities the change in breakdown voltage between 77°K and 473°K is 2 volts. This is somewhat larger than expected based on Tyagi's measurements.[32] At a diode current of 500 ma the difference in terminal voltage increases to 4 volts. If we attribute 2 volts to the change in breakdown voltage the additional 2 volts would have to be assigned to changes in mobility in the base region, a value which is not unrealistic at the high base doping levels.

The space charge resistance appears to be negligible in these diodes. First of all the voltage-current plot is a straight line over the entire current range. Secondly, the junction is a linear graded type with the electric field a parabolic function of distance. This situation is somewhere between that of an abrupt junction and a p-i-n junction; hence, a value close to zero is not unrealistic. Finally, the current density at 1 amp is an order of magnitude too low to affect the space charge at the surface of the base region. The high doping level in effect pins the edge of the depletion layer and prevents it from expanding into the base region.

In summary, it appears that the diode resistance in the avalanche breakdown region is accounted for mostly by the series base resistance. The "thermal resistance" is much smaller and is divided equally between changes in ionization coefficients and changes in mobility. Space charge resistance can be neglected.

We are now in a position to describe quantitatively the power

density throughout the diode. Consider a reverse biased diode at room temperature carrying 500 ma. About 12 volts appears across the base region, 6.5 volts across the depletion layer roughly 0.1 μm wide, and about 0.5 volts across the emitter material. If the power is dissipated in a 2 μm layer the power densities in the emitter, depletion layer, and base regions are 3×10^{-4}, 3×10^{-1}, and 6×10^{-3} watts/μm^3, respectively.

The current, power, and temperature distribution are illustrated in Figure 66. As the temperature rises the power distribution changes somewhat due to the thermal resistance. The changes also depend on the pulse generator characteristics. Eventually the temperature reaches a critical point where a constriction occurs either at the junction or in the base material. At a still later time second breakdown occurs (at about 600 nsec at these power levels).

Estimates of the junction temperature when constriction occurs are difficult to make for diodes in the reverse bias mode. Temperatures based on solutions to the linear heat equation are subject to the limitations pointed out in Section II. The solutions by numerical techniques in that section are inapplicable because they are based on one-dimensional problems. It is clear from Figure 66 that two and preferably three dimensional models are required.

Junction temperatures were estimated from extrapolated steady state threshold power versus ambient temperature curves. The results showed that second breakdown does not occur until a temperature of about 1100°C is reached. This is unexpectedly high, but other evidence presented later also attests to such high temperatures.

When these same diodes are forward biased a completely different

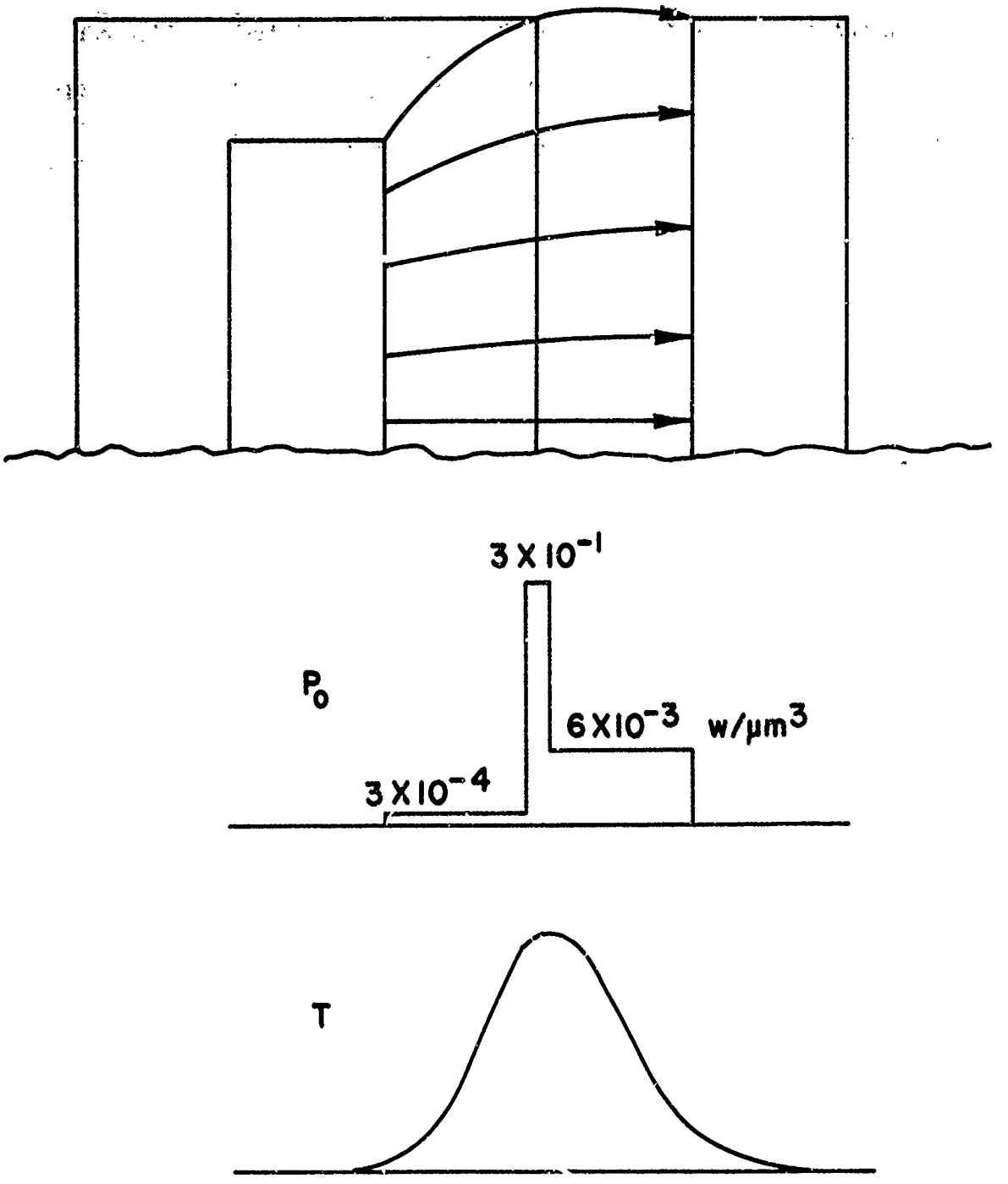


Fig. 66. Current, power, and temperature distribution in Part A for the reverse bias case. The power density distribution is calculated for a current of 500 ma.

situation holds. In the first place the voltage across the junction can be ignored. Secondly, conduction is not restricted solely to the active edge of the junction but includes much more junction area. Thirdly, the incremental diode resistance at currents greater than a few hundred milliamps is only about 4.5 ohms, indicating a high degree of conductivity modulation by injected crrriers. All of this means a relatively uniform distribution of power over a fairly broad area of the diode, and that the computer solutions in Section II are valid over short time intervals where lateral heat flow can be ignored.

Figure 29 shows that second breakdown occurs in 100 nsec at a pulse power of 180 watts in the forward bias condition. If the forward bias curve is replotted to indicate threshold power for constriction formation (assumed to occur when the pulse voltage reaches its maximum value) the power level is 150 watts at 100 nsec. Taking the power dissipating region as extending over most of the emitter and base regions between centerlines of the contacts, and assuming a 2 μm depth the power density is 3.8×10^6 watts/cm^2-μm. Figure 17 shows that the temperature rises to about 1300°C when the constriction forms.

The threshold curve for constriction formation should coincide with one of the curves in Figures 17-19 at short pulse lengths. The best fit occurred with the 1200°C curve in Figure 18 for t up to 2 μsec. The confidence in such curve fitting is limited by the small amount of data for the forward biased direction, but the result reinforces the contention that the peak temperatures are very high at the onset of current constriction.

The temperature rise in the diodes of Part B, the special test device, can be estimated by the same techniques. In the reverse bias mode

the temperature reaches about 1000°C before a constriction forms. In the forward bias direction the critical temperature is about 1100°C.

The power density distribution in the transistors and other three layer structures is complicated by the interaction between junctions, and the multiple conduction paths. Since the distribution is non-uniform the temperature rise cannot be determined from the curves of Section II. A tabulation of threshold power and power per unit area allows for some interesting comparisons and is presented in Table III.

The numbers in the Device column give the pin connections for different configurations or different geometries. The areas are calculated in most cases from the distance between contact centerlines and a distance slightly greater than the length of the active edge of the junction. Some question can be rightfully raised about the significance of the area calculated in this manner. The length of the active periphery is more important in some cases. On the other hand base width has been shown to be an important factor as has junction depth. When two junctions interact the concept of an active perimeter diminishes in importance. For these reasons the device area was chosen as the normalizing factor even though it can be misleading in some cases. Threshold power was taken directly from the experimental curves, extrapolated to steady-state conditions and to 0.1 μsec where necessary.

With due regard for the vagaries introduced by the device area it is still possible to discern trends in the figures of Table III. In the first place the largest devices tend to have the lowest power/area ratio even though the total power may be greater. This is expected in a steady-state situation because the thermal spreading resistance is a nonlinear function of area. However, even at 0.1 μsec the disparity has not diminished. It is

TABLE III

THRESHOLD POWER AND POWER PER UNIT AREA

Device	Area(μm^2)	Power (watts) and Power/Area (10^6 watts/cm^2)					
		0.1 μsec		10 μsec		DC	
		P	P/A	P	P/A	P	P/A
Part A	1,800	32	1.8	3.7	.21	2.8	.16
Part B Diode	12,000	180	1.5	19	.16	6.5	.054
E-B(1-5)	10,000	90	.90	14	.14	6.5	.065
E-B(1-3)	9,400	53	.57	11	.12	6.0	.064
C-B(2-3)	28,000	190	.68	16	.057	4.5	.016
C-B(4-5)	28,000	220	.79	19	.068	5.5	.020
C-E	20,000	900	4.5	42	.21	8.0	.04
Part C(1-2)	2,300	130	5.7	5.7	.25	3.2	.14
(9-7)	1,200	46	3.8	5.3	.44	3.2	.27
(3-2)	900	22	2.4	3.0	.33	2.2	.25
Part D	900	110	12.2	4.2	.47	2.7	.30
Part E	1,000	36	3.6	3.5	.35	2.0	.20

difficult to see why this should depend on device area, hence can only be attributed to differences in device "hardness."

The other conclusion derived from Table III is that glassivated and dielectrically isolated devices have higher threshold power/area. Whether this virtue is a generic property of these two technologies cannot be firmly established on the basis of these limited results. The uniformity and consistency of the data in this regard are too strong to be ignored, but arguments presented below suggest that the cause is due to construction features and not to advantages inherent in any one technology.

The question remains, why is there a twenty-fold difference in power/area among the circuits tested? Why is one circuit harder than another? The answer is contained in the two general concepts of power distribution and critical temperature.

The more uniform the power density, other things being equal, the greater the power handling capability. This is the main reason that a diode requires more power in the forward direction to reach second breakdown. In the reverse direction the high power density at the junction produces a local peak in the temperature distribution (Figure 66). If the peak were ironed out it would take longer to reach the critical temperature.

This is also the reason that a low voltage junction is harder than a high voltage junction. Note the low threshold power per unit area for collector-base junctions. Again, the power density is much greater in the depletion layer of the high voltage junction. For a given device area any modification which smooths out the power distribution will lead to greater power handling capability. Practices which increase the parasitic resistance, decrease breakdown voltages, increase active periphery, increase

junction depths, increase contact areas, etc., will all lead to increased threshold power.

The second important concept is that of a critical temperature. The device which has the highest critical temperature, other things being equal, will have the highest threshold power. Any fabrication technique which increases the critical temperature will increase the threshold power. In this category should also be included heat sinking improvements which restrain the rise in temperature.

Within this framework how can we explain the relative hardness of certain circuits? All of the circuits have roughly the same breakdown voltages, sheet resistivities, junction depths, etc. The only distinguishing feature of the hard circuits is the possibility of multiple current paths, in particular conduction via collector material or a buried layer as well as along the surface. The low threshold circuits, Parts A and B, have generally only a surface conduction path. Device C, C-B junction, has multiple paths. The low threshold can be explained however by the high breakdown voltage and hence high power density. The collector-emitter connection can also conduct along multiple paths. Significantly, this connection can be classified as hard for short pulses. Thus there is a certain correlation between hardness and current flow along subsurface paths.

What about the effects of heat sinking by glass and oxide layers? The influence of surface layers of thermal oxide or deposited glass does not have a significant effect on the peak temperature. This was demonstrated in Section II and has been pointed out earlier.[12] The oxide isolation in dielectrically isolated circuits has been shown to have a negligible effect on peak temperature for pulses up to 10 μsec![12] The present results indicate

an insignificant effect for up to 10 msec. The thermal resistance of the oxide layer is relatively small because of the large area to thickness ratio; furthermore, its effect would be to decrease rather than increase the threshold power.

Finally, one should consider the possibility that the critical temperature of the glassivated and dielectrically isolated circuits is significantly higher for some reason associated with the fabrication process. The critical temperature of Parts A and B was estimated to be in the range 1000-1200°C. The maximum increase in critical temperature could only be 200-400°C. The threshold curves for Device A, Figure 29, show that this increase only accounts for a threefold increase in power levels, even in a temperature range where the thermal conductivity is changing rapidly.

Thus it appears that differences in circuit performance are due primarily to differences in component area and differences in the uniformity of the power distribution.

SECTION VI

CONCLUSION

This research program has resulted in a number of interesting observations which are summarized in this section. Some of these are new, others corroborate findings of other workers. Some are based on firm information, others must be regarded as tentative. They are listed here in summary form without elaboration.

A. Heat Conduction

1. The thermal conductivity of silicon varies so drastically with temperature that the assumption of a constant value leads to large errors in heat flow calculations unless the temperature range is small.

2. Surface layers of thermal oxide or deposited glass have little heat sinking capability. Furthermore layers thicker than 1 μm prevent heat sinking by deposited metal films for pulse lengths of 1 μsec or less.

3. Plated heat sinks adjacent to heat producing junctions have little effect for pulses less than 10 nsec.

B. Second Breakdown in Diodes

1. The second breakdown model of Budenstein et al., applied to the integrated circuits tested here. In the reverse direction a current constriction forms at the junction, followed by growth of a current filament into the high resistivity material. When the filament reaches the base contact second breakdown occurs. A melted column is produced and permanent damage occurs.

2. In the forward bias direction the filament forms in the high resistivity region and grows to entirely span the region. When this is completed second breakdown occurs.

3. Temperatures above 1000°C are reached before constriction occurs.

4. Variations in threshold power for a given part are caused by variations in the series bulk resistance.

5. The threshold level can be changed by simple changes in construction. If the base and collector regions are shorted together to form the anode some of the current flows along a subsurface path. The broader distribution of power produces a harder junction.

C. Second Breakdown in Transistors

1. Emitter-base breakdown is similar to ordinary p-n junction breakdown.

2. In collector-base breakdown a dual path for current flow exists. At short pulse lengths failure occurs along the surface with degradation of the emitter-base junction. At long pulse lengths damage occurs beneath the surface and the collector base junction is degraded.

3. The threshold power per unit is smallest for collector-base pulsing because of the high power density at the junction.

4. Collector-emitter breakdown occurs in two stages. The first stage is nondestructive and takes place when a current constriction elongates from the collector junction to the collector contact. In the second stage the filament grows through the base region and to the emitter contact. Permanent damage then results.

5. The collector-emitter configuration is the hardest, both in terms of total power and power per unit area.

D. Areas for Further Study

1. A three-dimensional analysis of the heat flow problem would be extremely valuable in analyzing second breakdown phenomenon. The lack of

definitive temperature distributions in reverse biased diodes was a handicap in this study.

2. Further testing and analysis of transistors and certain diode structures is needed to verify details of the models proposed here.

REFERENCES

1. V. E. Holt, R. J. Grosh, and R. Geynet, "Evaluation of the Net Radiant Heat Transfer Between Specularly Reflecting Plates," Bell System Tech. J., Vol. 41, pp. 1865-1874, Nov. 1972.

2. H. S. Carslow and J. C. Jaeger, Conduction of Heat in Solids, Second Edition, Oxford University Press, London, 1959.

3. E. Mosekilde, P. A. Lebwohl, and D. G. Carlson, "Transient Heat Flow From an Epitaxial Layer into the Substrate," Proc. IEEE, Vol. 59, pp. 1030-1032, June 1971.

4. R. D. Lindsted and R. J. Surty, "Steady-State Junction Temperatures of Semiconductor Chips," IEEE Trans. Electron Devices, Vol. ED-19, pp. 41-44, Jan. 1972.

5. D. P. Kennedy, "Spreading Resistance in Cylindrical Semiconductor Devices," J. App. Physics, Vol. 31, pp. 1490-1497, Aug. 1960.

6. D. J. Fleming, "Thermal Breakdown Delay Time in Silicon p-n Junctions," IEEE Trans. Electron Devices, Vol. ED-18, pp. 94-97, Feb. 1971.

7. C. Kittel, Introduction to Solid State Physics, Fourth Edition, John-Wiley and Sons, New York, 1971.

8. J. M. Ziman, Electrons and Phonons, Oxford University Press, London, 1960.

9. C. J. Glassbrenner and G. A. Slack, "Thermal Conductivity of Silicon and Germanium from 3°K to the Melting Point," Phys. Rev., Vol. 134, pp. A1058-A1069, 18 May 1964.

10. G. A. Slack, "Thermal Conductivity of Pure and Impure Silicon, Silicon Carbide, and Diamond," J. App. Phys., Vol. 35, pp. 3460-3466, Dec. 1964.

11. J. S. Smith, "High Current Transient Induced Junction Shorts," 9th Annual Proceedings on Reliability Physics 1971, pp. 163-171.

12. J. S. Smith, "Pulse Power Testing of Microcircuits," Rome Air Development Center Technical Report, RADC-TR-71-59, Oct. 1971, AD 733 342.

13. J. Todd, Survey of Numerical Analysis, McGraw Hill, N. Y., 1962.

14. R. D. Richtmeyer, Difference Methods for Initial-Value Problems, Interscience Publishers, New York, 1957.

15. C. G. Thornton and C. D. Simmons, "A New High Current Mode of Transistor Operation," IRE Trans. Electron Devices, Vol. ED-5, pp. 6-10, Jan. 1958.

16. J. Tauc and A. Abraham, "Thermal Breakdown in Silicon p-n Junctions," Phys. Rev., Vol. 108, pp. 936-937, Nov. 15, 1957.

17. H. A. Schafft, "Second Breakdown - A Comprehensive Review," Proc. IEEE, Vol. 55, pp. 1272-1288, Aug. 1967.

18. G. M. Ford, "Collector to Emitter Breakdown Related to Thermal Runaway in Homogeneous Base Germanium Power Transistors," Solid State Design, Vol. 4, pp. 29-36, June 1963.

19. H. Melchior and M.J.O. Strutt, "Secondary Breakdown in Transistors," Proc. IEEE, Vol. 52, pp. 439-440, April 1964.

20. A. C. English and H. M. Power, "Mesoplasma Breakdown in Silicon Junctions," Proc. IEEE, Vol. 51, pp. 500-501, March 1963.

21. H. C. Bowers, "Space-Charge-Induced Negative Resistance in Avalanche Diodes," IEEE Trans. Electron Devices, Vol. ED-15, pp. 343-350, June 1968.

22. M. W. Muller and H. Guckel, "Negative Resistance and Filamentary Currents in Avalanching Silicon p^+-i-n^+ Junctions," IEEE Trans. Electron Devices, Vol. ED-15, pp. 560-568, Aug. 1968.

23. P. L. Hower and V.G.K. Reddi, "Avalanche Injection and Second Breakdown in Transistors," IEEE Trans. Electron Devices, Vol. ED-17, pp. 320-335, April 1970.

24. D. M. Tasca, "Pulse Power Failure Modes in Semiconductors," IEEE Trans. Nuc. Sci., Vol. NS-17, pp. 364-372, Dec. 1970.

25. K. L. Chiang and P. O. Lauritzen, "Thermal Instability in Very Small p-n Junctions," IEEE Trans. Electron Devices, Vol. ED-17, pp. 782-787, Sept. 1970.

26. C. Wang, "Temperature Effects in Semiconductor Avalanches," Ph.D. thesis, Cornell University, June 1970.

27. R. A. Sunshine and M. A. Lampert, "Second-Breakdown Phenomena in Avalanching Silicon-on-Sapphire Diodes," IEEE Trans. Electron Devices, Vol. ED-19, pp. 873-885, July 1972. Also R. A. Sunshine, "Avalanching and Second Breakdown in Silicon-on-Sapphire Diodes," University Microfilms, 1971.

28. D. J. Dumin, "Emission of Visible Radiation from Extended Plasmas in Silicon Diodes During Second Breakdown," IEEE Trans. Electron Devices, Vol. ED-16, pp. 479-485, May 1969.

29. P. P. Budenstein, D. H. Pontius, and W. B. Smith, "Second Breakdown and Damage in Semiconductor Junction Devices," Final Report, Contract No. DAAH 01-71-C-0708, U. S. Army Missile Command, AD740226.

30. R. Brown, J. D. Holder, and V. W. Ruwe, "Statistical Component Damage Study," Conference Proceedings on Component Degradation From Transient Inputs, U. S. Army Mobility Equipment Research and Development Center, Fort Belvoir, Virginia.

31. D. M. Tasca and J. C. Peden, "Pulsed Power Failure Modes in Semiconductor Devices," Conference Proceedings on Component Degradation From Transient Inputs, U. S. Army Mobility Equipment Research and Development Center, Fort Belvoir, Virginia.

32. M. Singh Tyagi, "Zener and Avalanche Breakdown in Silicon Alloyed p-n Junctions - I" Solid State Electronics, Vol. 11, pp. 99-115, 1968. Also "Zener and Avalanche Breakdown in Silicon Alloyed p-n Junctions - II," ibid., pp. 117-128.

APPENDIX

The computer program described here is written in Fortran for use on a Honeywell 635 time-sharing system. Most of the calculations in Section II were performed with this program.

As shown here the program solves transient heat flow problems containing up to 20 nodes of arbitrary sizes consisting of silicon, silicon dioxide, or aluminum in any order. The initial temperature can have any value, but must be uniform. Power can be supplied in any region but cannot be a function of time. The last node is fixed at the ambient temperature.

To use the program the operator divides the device into regions containing any number of nodes but only one kind of material. For each region the dimension of the region in microns, the number of nodes, the power density in watts/cm^2-μm, and the kind of material must be specified (1 for aluminum, 2 for silicon dioxide, 3 for silicon). Also the ambient temperature, time interval between iterations, final time for the problem, and the number of iterations between printout must be typed in. The computer then lists the nodes, their location, the time and the node temperatures.

Statement numbers 150-470 read in all the input data. Statements 480-640 calculate the size and location of each node. A portion of the printout is handled by statements 650-720.

Statements 730-1770 are a tabulation of the thermal conductivity and specific heat of aluminum, silicon dioxide, and silicon from 0 to 1700°K at 50 degree intervals. A straight line interpolation is used at intermediate temperatures.

The heat flow equations are set up in statements 1800-2230. New coefficients are calculated at the beginning of each iteration. The

equations are solved in statements 2240-2330.

An example of a problem solution is included. The device is divided into 11 regions. The first region is an aluminum layer 1 μm thick containing 1 node. The second region is silicon dioxide, 0.5 μm thick, with 1 node. The third region is silicon, 3 μm thick, with 3 nodes. The fourth region represents the depletion layer of a junction and contains the source of heat. This region is 1 μm thick, has 1 node, and has a power density of 6×10^5 watts/cm^2-μm. This is equivalent to about 50 watts in a 1 μm layer of a device measuring 70 x 100 μm.

Regions 5 through 10 are all silicon with gradually increasing node size to provide a smooth transition to a relatively thick substrate. For region 11 a group of zeros is typed in to terminate the input data.

The ambient temperature is 27°C. The time interval is 10 nsec with a final time of 2.1 μsec.

```
10******THAP - TRANSIENT HEAT ANALYSIS PROGRAM. MAXIMUM NUMBER OF
20******NODES = 20. NODE 1 HAS ZERO HEAT TRANSFER OUT OF THE SURFACE.
30******LAST NODE CONSTRAINED TO AMBIENT TEMPERATURE.
40       DIMENSION SILICONK(35), SILICONC(35), SIO2K(35),
50&      SIO2C(35), ALK(35), ALC(35), X(20), DX(20), Q(20),
60&      T(20), NODES(20), XR(20), QR(20), NR(20), MR(20),
70&      MAT(20), R(20), RHOCDD(20), A(20,21), E(20),
80&      F(20)
90       PRINT 99
100    99 FORMAT( ////"            #####  THAP - TRANSIENT
110& HEAT ANALYSIS PROGRAM  #####")
120       PRINT 100
130   100 FORMAT( ///"            DIMENSIONS, NODES, HEAT IN
140&PUT, MATERIAL"/)
150******TYPE IN DIMENSIONS OF EACH REGION IN MICRONS, NUMBER OF
160******NODES IN THAT REGION, POWER DENSITY IN WATTS/CM2/MICRON,
170******AND TYPE OF MATERIAL; ALUMINUM=1, SILICON DIOXIDE=2,
180******AND SILICON=3.
190       DO 102 NI = 1,20
200       PRINT 103, NI
210   103 FORMAT( " REGION "I2)
220       READ: XR(NI),NR(NI),QR(NI),MR(NI)
230       NII = NI-1
240       N = N + NR(NI)
250       IF(XR(NI)) 102,104,102
260   102 CONTINUE
270   104 CONTINUE
280       M = N + 1
290       L = N - 1
300       PRINT 110
310   110 FORMAT( //" AMBIENT TEMPERATURE")
320******TYPE IN AMBIENT TEMPERATURE IN DEGREES CENTIGRADE
330       READ: TA
340       DO 111 IN = 1,20
350   111 T(IN) = TA
360       PRINT 150
370   150 FORMAT( //" TIME INTERVAL")
380******TYPE IN TIME INTERVAL(ALSO INITIAL TIME) IN SECONDS,
390******FINAL TIME IN SECONDS, AND NUMBER OF TIME STEPS BEFORE
400******EACH PRINTOUT OCCURS
410       READ: DT
420       PRINT 151
430   151 FORMAT( " FINAL TIME")
440       READ: FT
450       PRINT 152
460   152 FORMAT( " STEPS BETWEEN PRINTOUT")
470       READ: NS
480******CALCULATE GEOMETRIC FACTORS
490       IB = 1
500       IC = NR(1)
```

```
      IC = NR(1)
      DO 125 IA = 1,NII
      XX = XR(IA)/NR(IA)
      QQ = QR(IA)
      MM = MR(IA)
      DO 126 I = IB,IC
      MAT(I) = MM
      DX(I) = XX
      Q(I) = QQ
  126 NODES(I) = I
      IB = IC+1
  125 IC = IC+NR(IA+1)
      X(1) = 0.5*DX(1)
      DO 127 I = 2,N
  127 X(I) = X(I-1)+0.5*(DX(I-1)+DX(I))
      PRINT 170
  170 FORMAT( ///"                                        SOLUTION ")
      PRINT 171, (NODES(I), I = 1,N)
  171 FORMAT( //" NODES     " 8I7/11X 8I7/12X 4I7)
      PRINT 172, (X(I), I = 1,N)
  172 FORMAT( " POSITION   "8F7.2/13X 8F7.2/14X 4F7.2)
      PRINT 173
  173 FORMAT( /"    TIME                                NODE TEMPERATURES")
******TABLE OF THERMAL PROPERTIES. K IS THERMAL CONDUCTIVITY
******IN WATTS/CM/DEGREE AND C IS SPECIFIC HEAT IN CALORIES
******/GRAM/DEGREE
      ALK(1) = 50.
      ALK(2) = 8.
      ALK(3) = 2.9
      ALK(4) = 2.5
      ALK(5) = 2.4
      ALK(6) = 2.39
      ALK(7) = 2.38
      ALK(8) = 2.38
      ALK(9) = 2.38
      ALK(10) = 2.38
      ALK(11) = 2.38
      DO 10 IND = 12,35
   10 ALK(IND) = ALK(IND-1) - 0.03
      ALC(1) = 0.
      ALC(2) = .04
      ALC(3) = .12
      ALC(4) = .165
      ALC(5) = .193
      ALC(6) = .207
      ALC(7) = .217
      ALC(8) = .225
      ALC(9) = .233
      DO 12 IND = 10,35
   12 ALC(IND) = ALC(IND-1) + .005
       SIO2K(1) = 0.
       DO 14 IND = 2,6
```

```
1020    14 SIO2K(IND) = SIO2K(IND-1) + .00272
1030       DO 16 IND = 7,35
1040    16 SIO2K(IND) = SIO2K(IND-1) + .000582
1050       SIO2C(1) = 0.
1060       SIO2C(2) = .034
1070       SIO2C(3) = .07
1080       SIO2C(4) = .097
1090       SIO2C(5) = .125
1100       SIO2C(6) = .15
1110       SIO2C(7) = .175
1120       SIO2C(8) = .19
1130       SIO2C(9) = .205
1140       SIO2C(10) = .22
1150       SIO2C(11) = .231
1160       SIO2C(12) = .242
1170       SIO2C(13) = .25
1180       SIO2C(14) = .256
1190       SIO2C(15) = .262
1200       SIO2C(16) = .267
1210       SIO2C(17) = .271
1220       SIO2C(18) = .275
1230       SIO2C(19) = .279
1240       SIO2C(20) = .282
1250       SIO2C(21) = .285
1260       SIO2C(22) = .288
1270       SIO2C(23) = .291
1280       DO 18 IND = 24,35
1290    18 SIO2C(IND) = SIO2C(IND-1) + .002
1300       SILICONK(1) = 0.
1310       SILICONK(2) = 24.
1320       SILICONK(3) = 8.8
1330       SILICONK(4) = 3.9
1340       SILICONK(5) = 2.6
1350       SILICONK(6) = 1.9
1360       SILICONK(7) = 1.5
1370       SILICONK(8) = 1.2
1380       SILICONK(9) = 1.0
1390       SILICONK(10) = .85
1400       SILICONK(11) = .75
1410       SILICONK(12) = .67
1420       SILICONK(13) = .61
1430       SILICONK(14) = .55
1440       SILICONK(15) = .50
1450       SILICONK(16) = .45
1460       SILICONK(17) = .42
1470       SILICONK(18) = .38
1480       SILICONK(19) = .35
1490       SILICONK(20) = .33
1500       SILICONK(21) = .31
1510       SILICONK(22) = .30
1520       SILICONK(23) = .29
```

```
1530       SILICONK(24) = .28
1540       SILICONK(25) = .28
1550       SILICONK(26) = .27
1560       SILICONK(27) = .27
1570       SILICONK(28) = .27
1580       SILICONK(29) = .26
1590       SILICONK(30) = .26
1600       SILICONK(31) = .26
1610       SILICONK(32) = .26
1620       SILICONK(33) = .26
1630       SILICONK(34) = .25
1640       SILICONK(35) = .25
1650       SILICONC(1) = 0.
1660       SILICONC(2) = .035
1670       SILICONC(3) = .070
1680       SILICONC(4) = .110
1690       SILICONC(5) = .130
1700       SILICONC(6) = .155
1710       SILICONC(7) = .175
1720       SILICONC(8) = .185
1730       SILICONC(9) = .195
1740       SILICONC(10) = .200
1750       SILICONC(11) = .205
1760       DO 21 KK = 12,35
1770    21 SILICONC(KK) = SILICONC(KK-1) + .002125
1780       TIME = 0.
1790       IS = FT/DT/NS
1800******SOLVE HEAT FLOW EQUATIONS
1810       DO 71 IT = 1,IS
1820       DO 70 ITT = 1,NS
1830******CALCULATE THERMAL PROPERTIES AT THE NODE TEMPERATURES
1840       DO 20 I = 1,N
1850       ENTER = (T(I) + 273.)/50. + 1.
1860       II = ENTER
1870       AA = II
1880       BB = ENTER - AA
1890       K = MAT(I)
1900       GO TO(62,61,60),K
1910    60 AK = SILICONK(II)+BB*(SILICONK(II+1)-SILICONK(II))
1920       RC = SILICONC(II)+BB*(SILICONC(II+1)-SILICONC(II))
1930******CONVERT SPECIFIC HEAT FROM CALORIES/GRAM/DEGREE TO
1940******JOULES/CM3/DEGREE
1950       RHOC = 4.184*2.328*RC
1960       GO TO 63
1970    61 AK = SIO2K(II)+BB*(SIO2K(II+1)-SIO2K(II))
1980       RC = SIO2C(II)+BB*(SIO2C(II+1)-SIO2C(II))
1990       RHOC = 4.184*2.20*RC
2000       GO TO 63
2010    62 AK = ALK(II)+BB*(ALK(II+1)-ALK(II))
2020       RC = ALC(II)+BB*(ALC(II+1)-ALC(II))
2030       RHOC = 4.184*2.707*RC
```

```
2040       GO TO 63
2050    63 R(I) = DX(I)/AK*1.E-4
2060    20 RHOCDD(I) = 1.E-4*RHOC*DX(I)/DT
2070******CALCULATE COEFFICIENTS IN EQUATIONS
2080       DO 30 I = 1,N
2090       IM = I-1
2100       IP = I+1
2110       IF(IM)31,31,32
2120    32 A(I,IM) = 2./(R(IM)+R(I))
2130       IF(IP-N)36,36,31
2140    36 A(I,I) = -2./(R(IM)+R(I))-2./(R(I)+R(IP))-
2150&      RHOCDD(I)
2160    31 A(I,M) = -DX(I)*Q(I)-RHOCDD(I)*T(I)
2170       IF(IP-N)34,34,35
2180    34 A(I,IP) = 2./(R(I)+R(IP))
2190    35 CONTINUE
2200    30 CONTINUE
2210       A(1,1) = -2./(R(1)+R(2))-RHOCDD(1)
2220       L = N-1
2230       A(N,N) = -2./(R(L)+R(N))-RHOCDD(N)
2240******SOLVE EQUATIONS USING RICHTMYER'S ALGORITHM
2250       F(1) = A(1,M)/A(1,1)
2260       E(1) = -A(1,2)/A(1,1)
2270       DO 50 J = 2,N
2280       F(J) = (A(J,M)-A(J,J-1)*F(J-1))/
2290&              (A(J,J)+A(J,J-1)*E(J-1))
2300    50 E(J) = -A(J,J+1)/(A(J,J)+A(J,J-1)*E(J-1))
2310       T(N-1) = TA*E(N-1) + F(N-1)
2320       DO 51 J = 2,L
2330    51 T(N-J) = E(N-J)*T(N-J+1) + F(N-J)
2340    TIME = TIME +DT
2350    70 CONTINUE
2360       PRINT 9, TIME, (T(JJ), JJ = 1,N)
2370     9 FORMAT( E10.3,2X 8F7.1/13X 8F7.1/14X 4F7.1)
2380    71 CONTINUE
2390    STOP
2400 END
```

```
        #####  THAP - TRANSIENT HEAT ANALYSIS PROGRAM  #####


        DIMENSIONS, NODES, HEAT INPUT, MATERIAL
REGION  1
= 1.,1,0.,1
REGION  2
= .5,1,0.,2
REGION  3
= 3.,3,0.,3
REGION  4
```

```
= 1.,1,6.E5,3
REGION  5
= 1.,1,0.,3
REGION  6
= 4.,2,0.,3
REGION  7
= 8.,2,0.,3
REGION  8
= 16.,2,0.,3
REGION  9
= 30.,3,0.,3
REGION 10
= 60.,4,0.,3
REGION 11
= 0,0,0,0


AMBIENT TEMPERATURE
= 27.  °C


TIME INTERVAL
= 10.E-9      seconds
FINAL TIME
= 2.1E-6
STEPS BETWEEN PRINTOUT
= 50
```

SOLUTION

NODES	1	2	3	4	5	6	7	8
	9	10	11	12	13	14	15	16
	17	18	19	20				
POSITION	0.50	1.25	2.00	3.00	4.00	5.00	6.00	7.50
	9.50	12.50	16.50	22.50	30.50	39.50	49.50	59.50
	72.00	87.00	102.00	117.00				

TIME	NODE TEMPERATURES							
0.500E-06	56.6	105.4	183.7	195.2	213.6	240.4	196.4	147.9
	106.2	68.1	45.0	31.0	27.6	27.1	27.0	27.0
	27.0	27.0	27.0	27.0				
0.100E-05	128.8	196.3	296.7	312.9	336.9	370.2	310.0	241.8
	179.2	119.5	77.3	45.2	31.9	27.9	27.1	27.0
	27.0	27.0	27.0	27.0				
0.150E-05	214.7	286.8	393.6	413.9	443.3	483.9	408.1	323.8
	244.8	167.3	109.4	53.2	39.7	30.3	27.7	27.1
	27.0	27.0	27.0	27.0				
0.200E-05	304.0	375.5	482.5	507.2	542.1	590.2	498.9	398.6
	305.0	212.5	140.7	82.1	49.3	34.3	29.0	27.5
	27.1	27.0	27.0	27.0				